

MATLAB EXPO

2021

将算法和模型部署到生产系统:嵌入式或企业系统

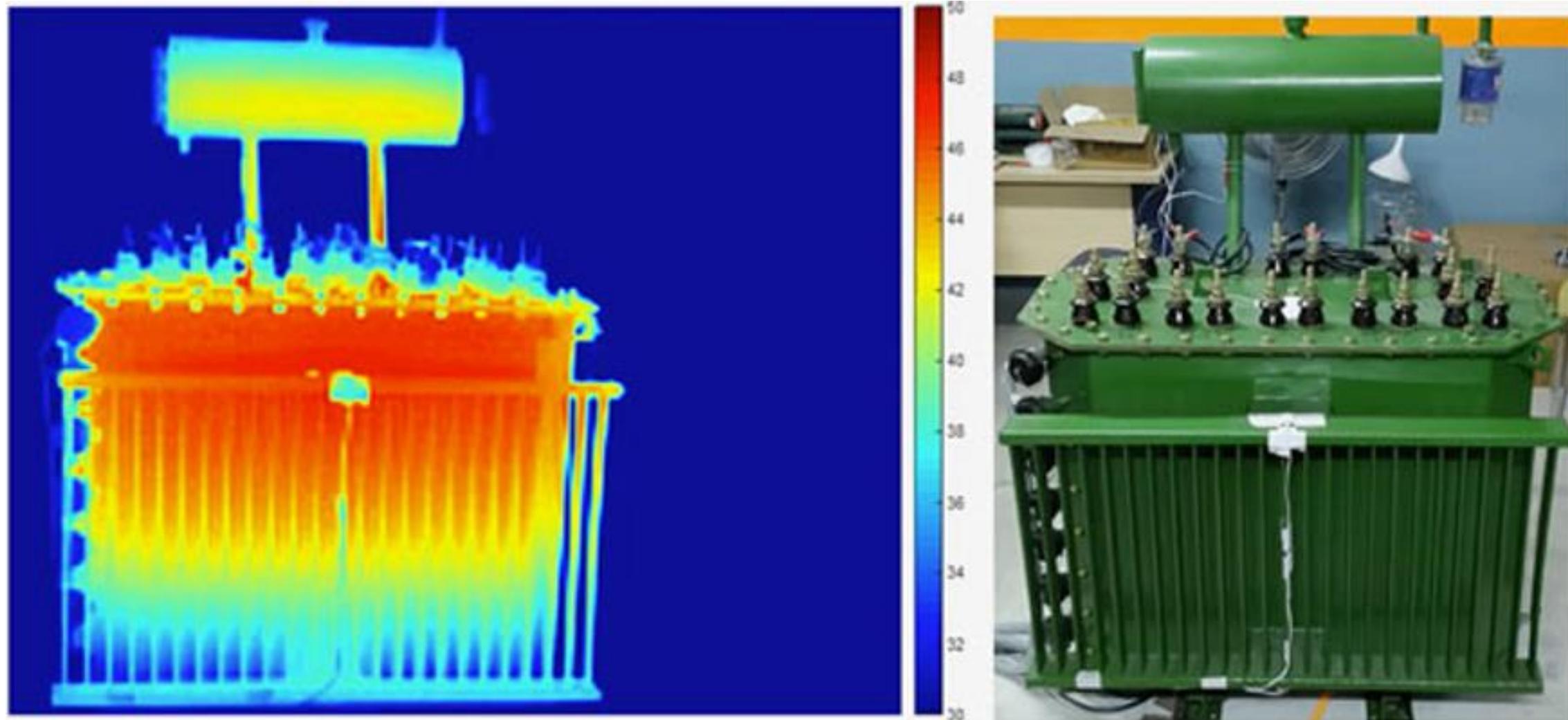


付玉敏, 上海电气集团中央研究院高级主管工程师

刘海伟, MathWorks 中国高级应用工程师

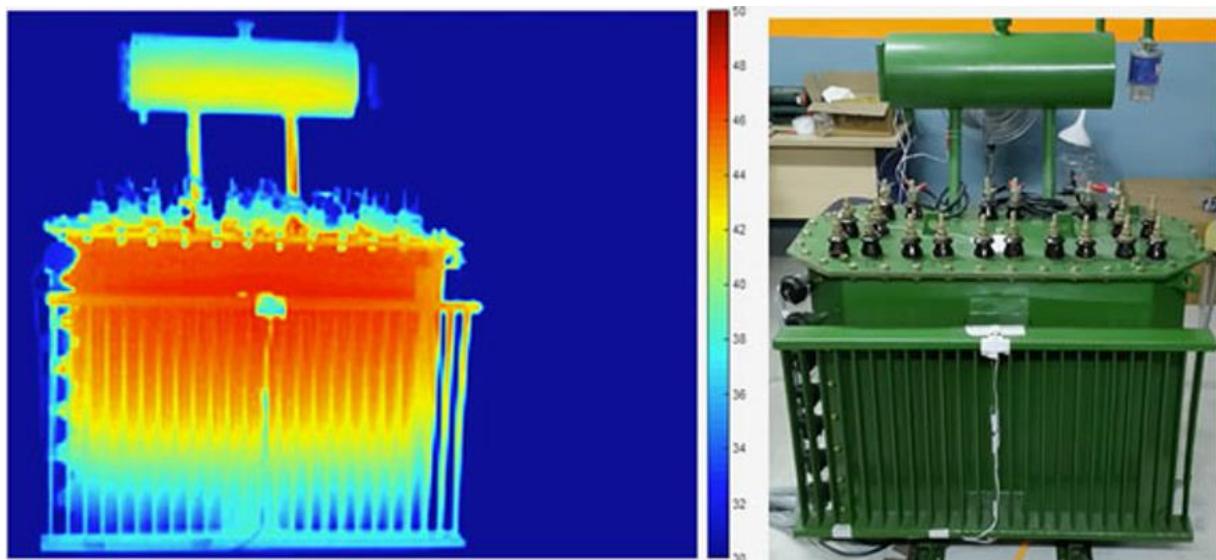


电网系统健康监测



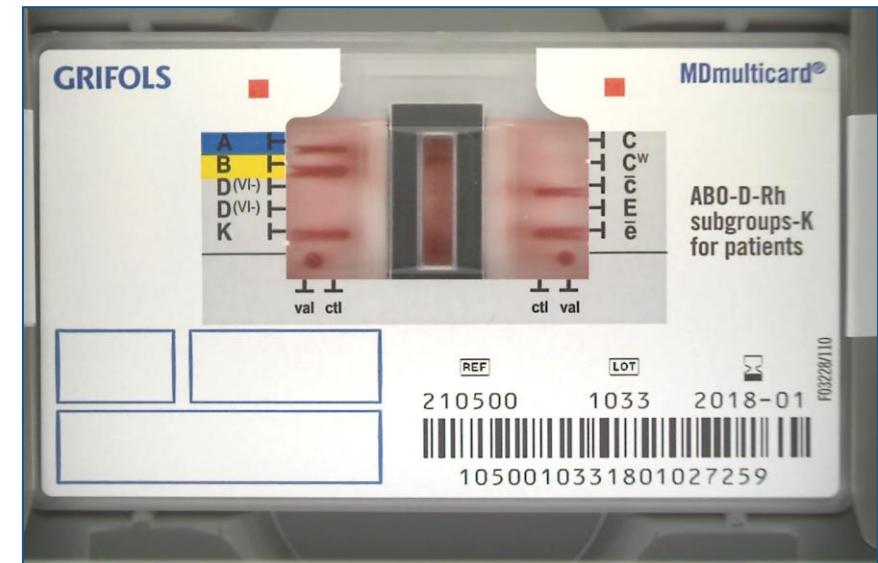
部署到嵌入式和企业系统

Enterprise



Health Monitoring of Distribution
Transformers **SIEMENS**

Embedded



Card to Classify Blood Type
IDNEO

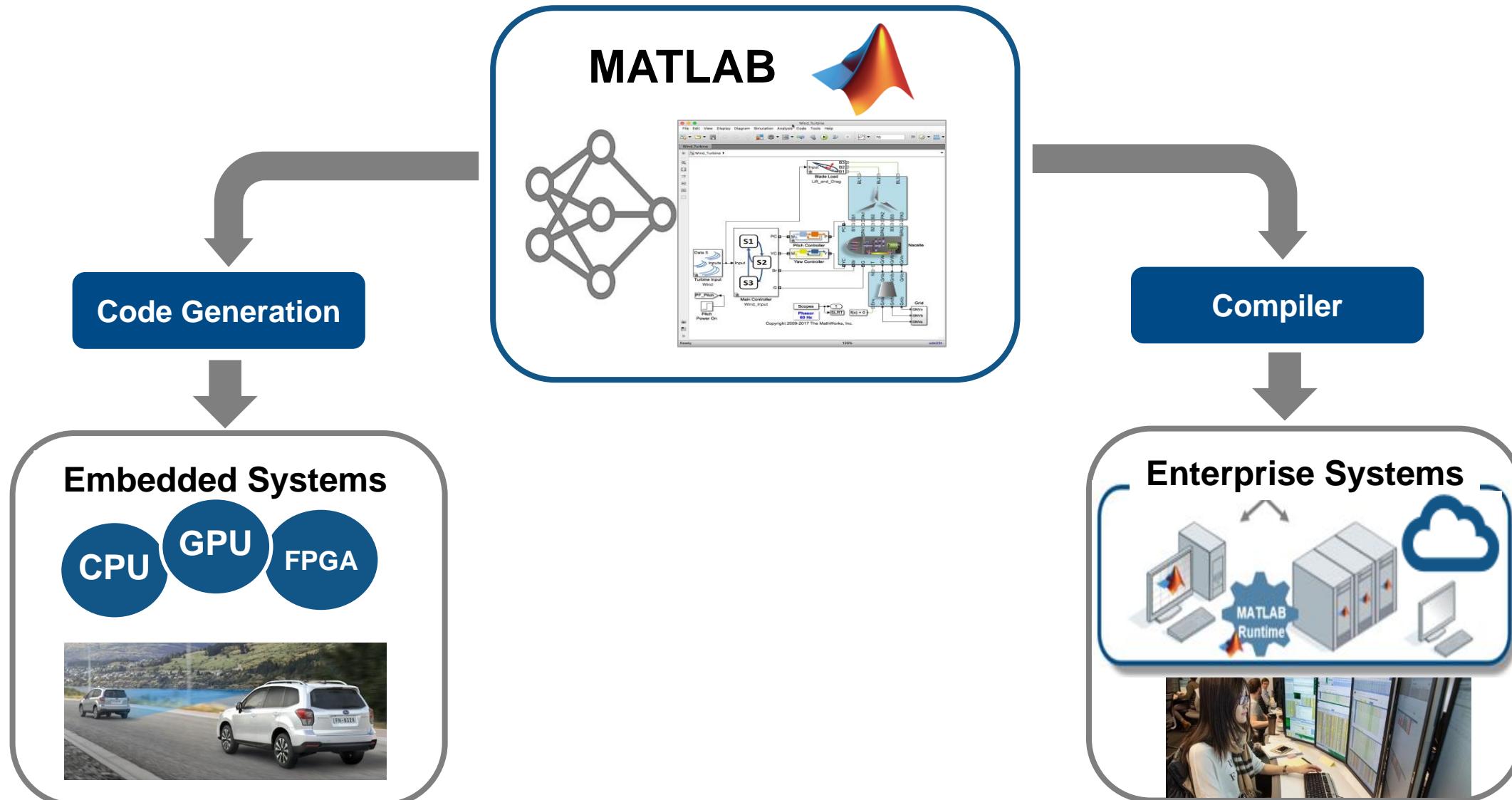
Agenda

算法和模型部署

三类特定问题:

1. 嵌入式硬件的约束
2. 环境和系统的持续变化
3. 扩展到企业生产环境

算法和模型部署的两个方向



声音场景识别算法的嵌入式部署

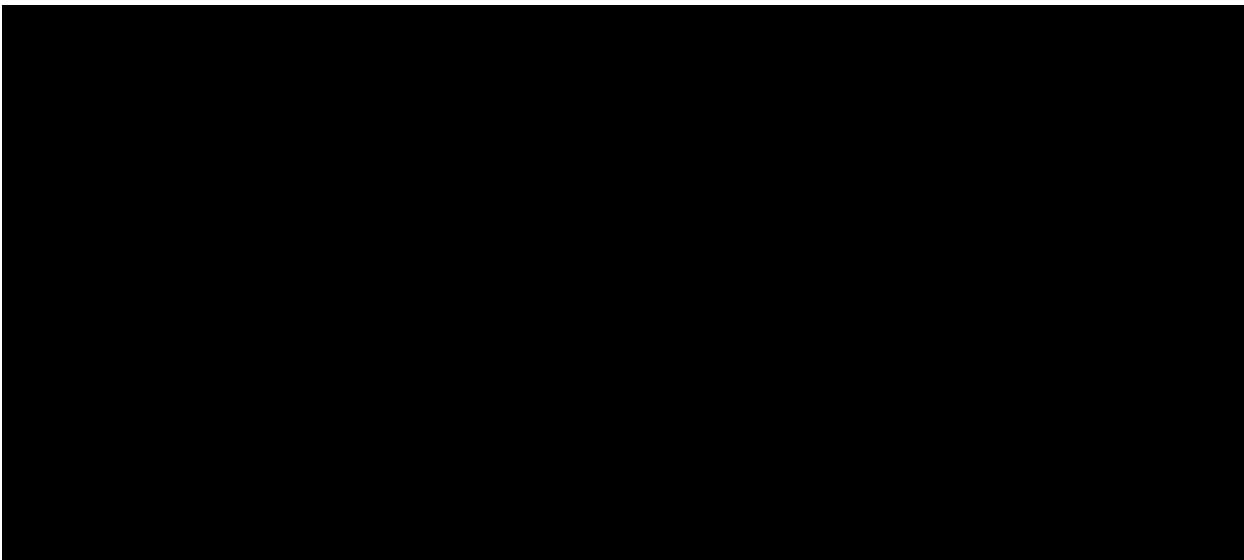
声音场景识别算法的嵌入式部署

SqueezeNet ~5MB
ResNet-50 ~100MB



Limited
resources

声音场景识别算法的嵌入式部署



SqueezeNet ~5MB
ResNet-50 ~100MB



Limited
resources

声音场景识别算法的嵌入式部署



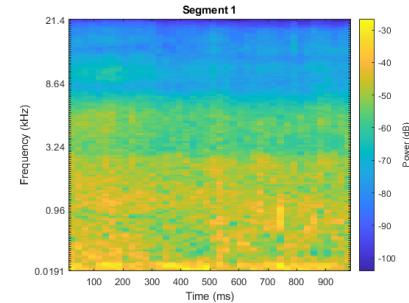
SqueezeNet ~5MB
ResNet-50 ~100MB



Limited
resources

声音场景识别算法的嵌入式部署

Reformat the data



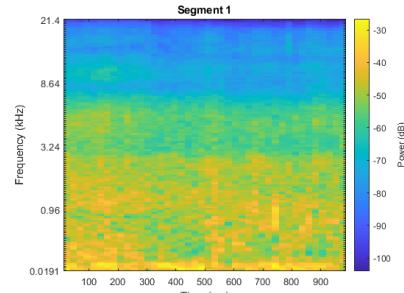
SqueezeNet ~5MB
ResNet-50 ~100MB



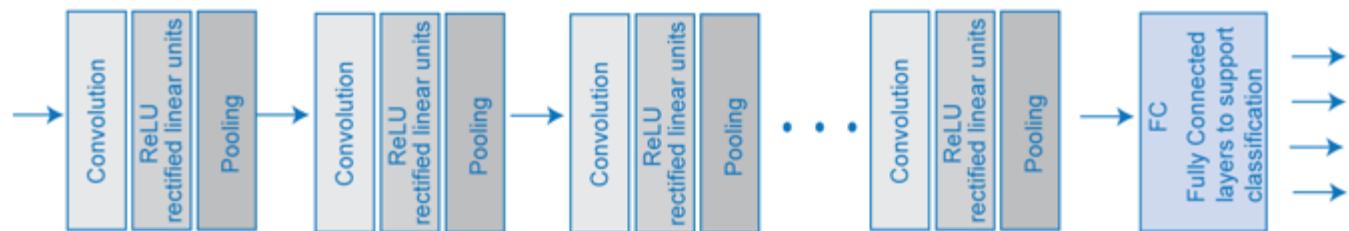
Limited
resources

声音场景识别算法的嵌入式部署

Reformat the data



Convolutional Neural Networks (**CNN**)



SqueezeNet ~5MB
ResNet-50 ~100MB



Limited resources

如何在嵌入式中使用模型

如何在嵌入式中使用模型



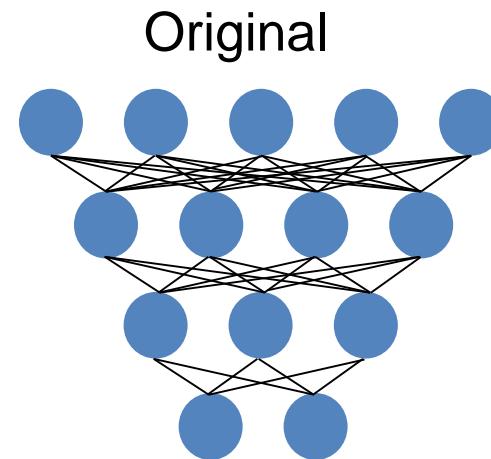
如何在嵌入式中使用模型



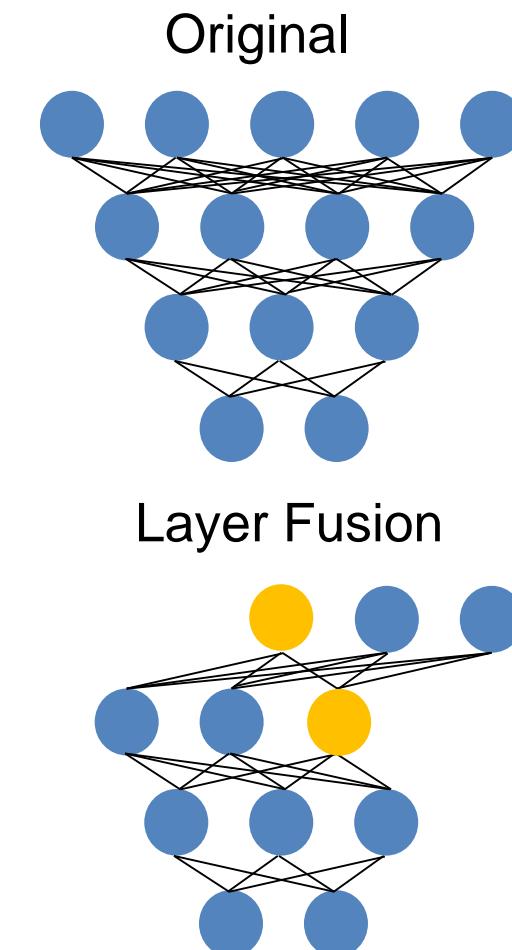
如何在嵌入式中使用模型



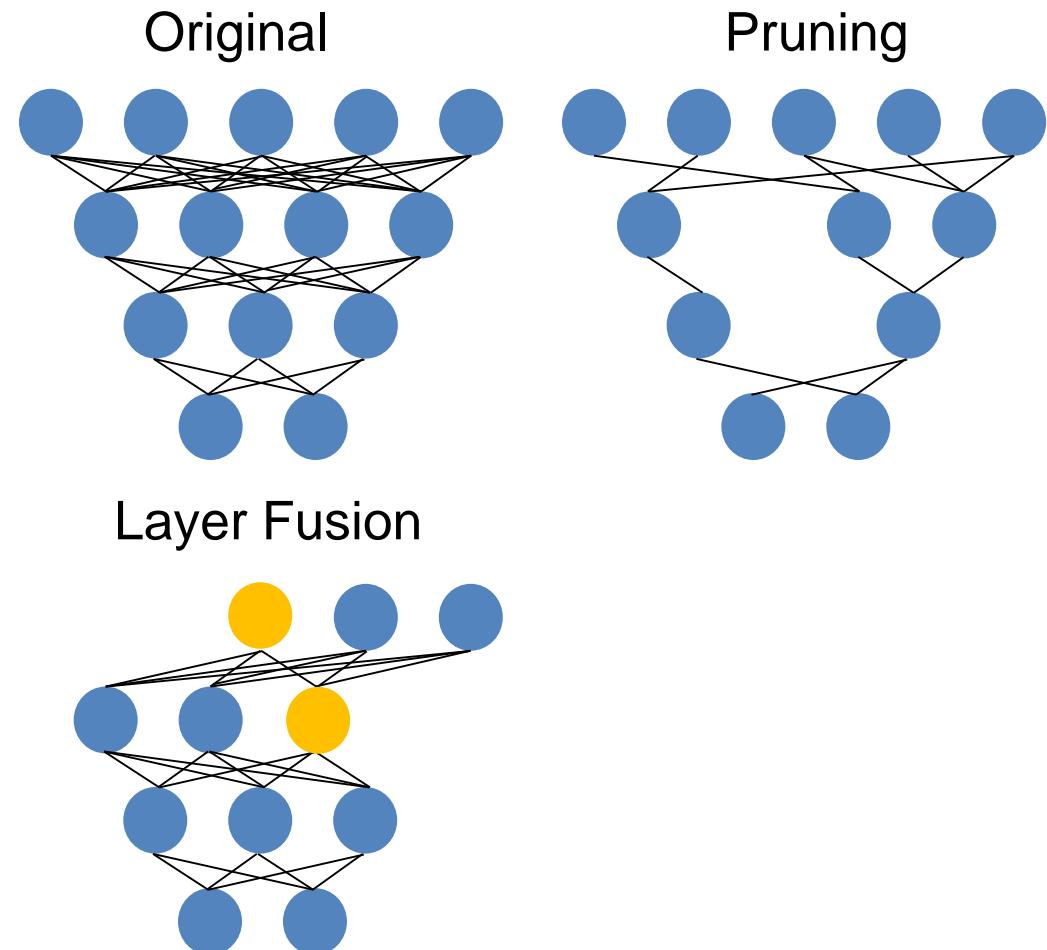
如何在嵌入式中使用模型



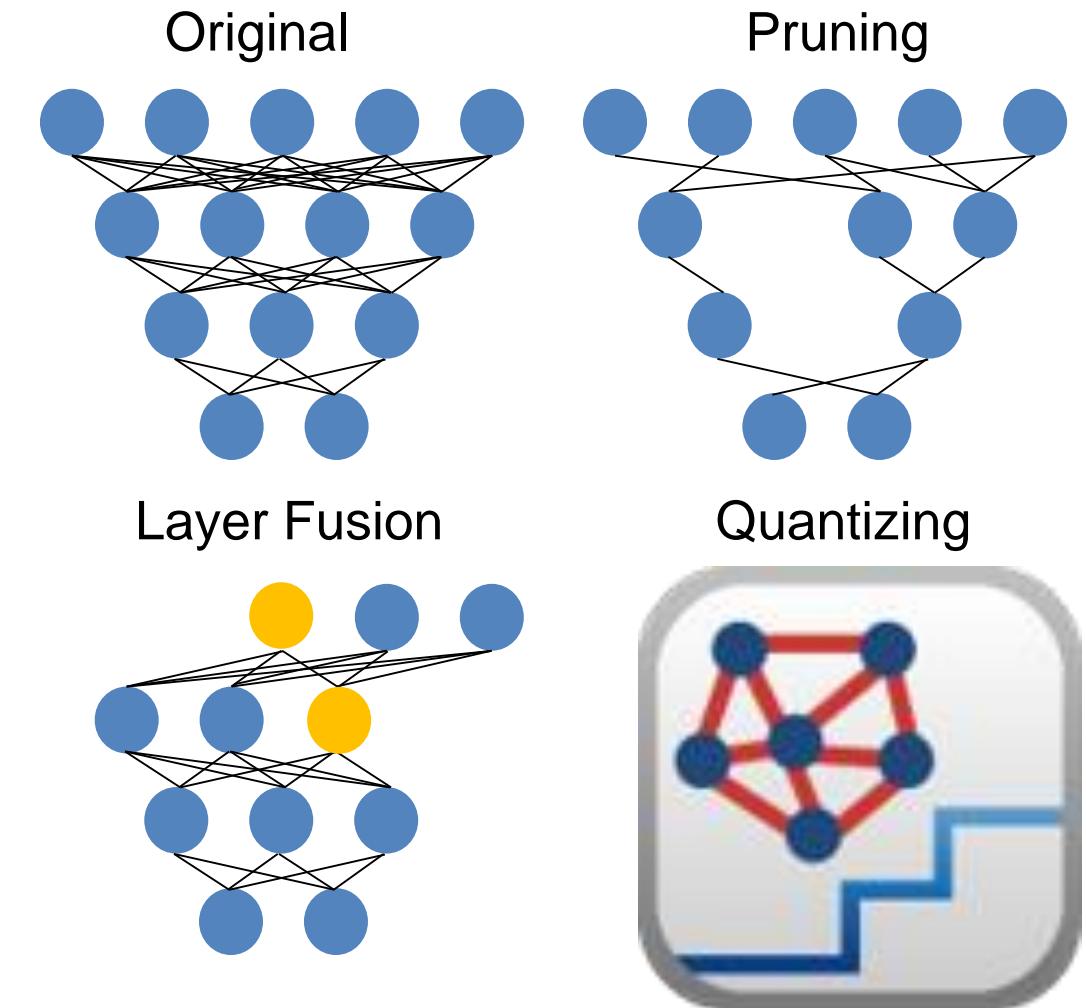
如何在嵌入式中使用模型



如何在嵌入式中使用模型



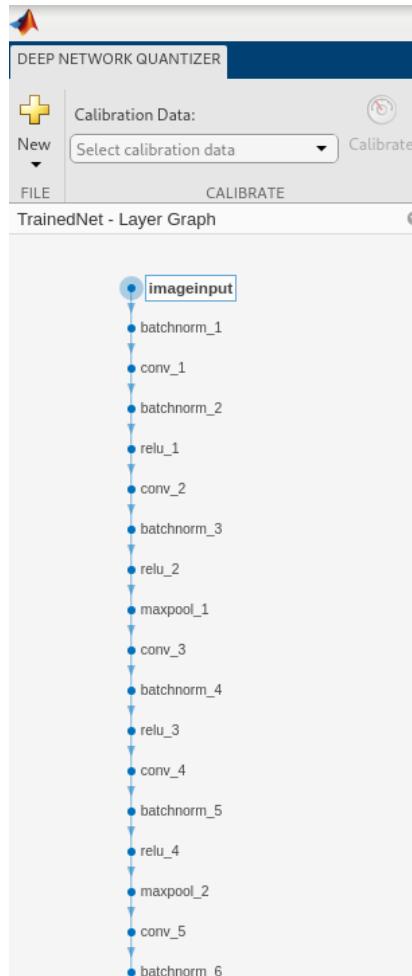
如何在嵌入式中使用模型



Deep Learning Quantization: 声音场景分类



Deep Learning Quantization: 声音场景分类

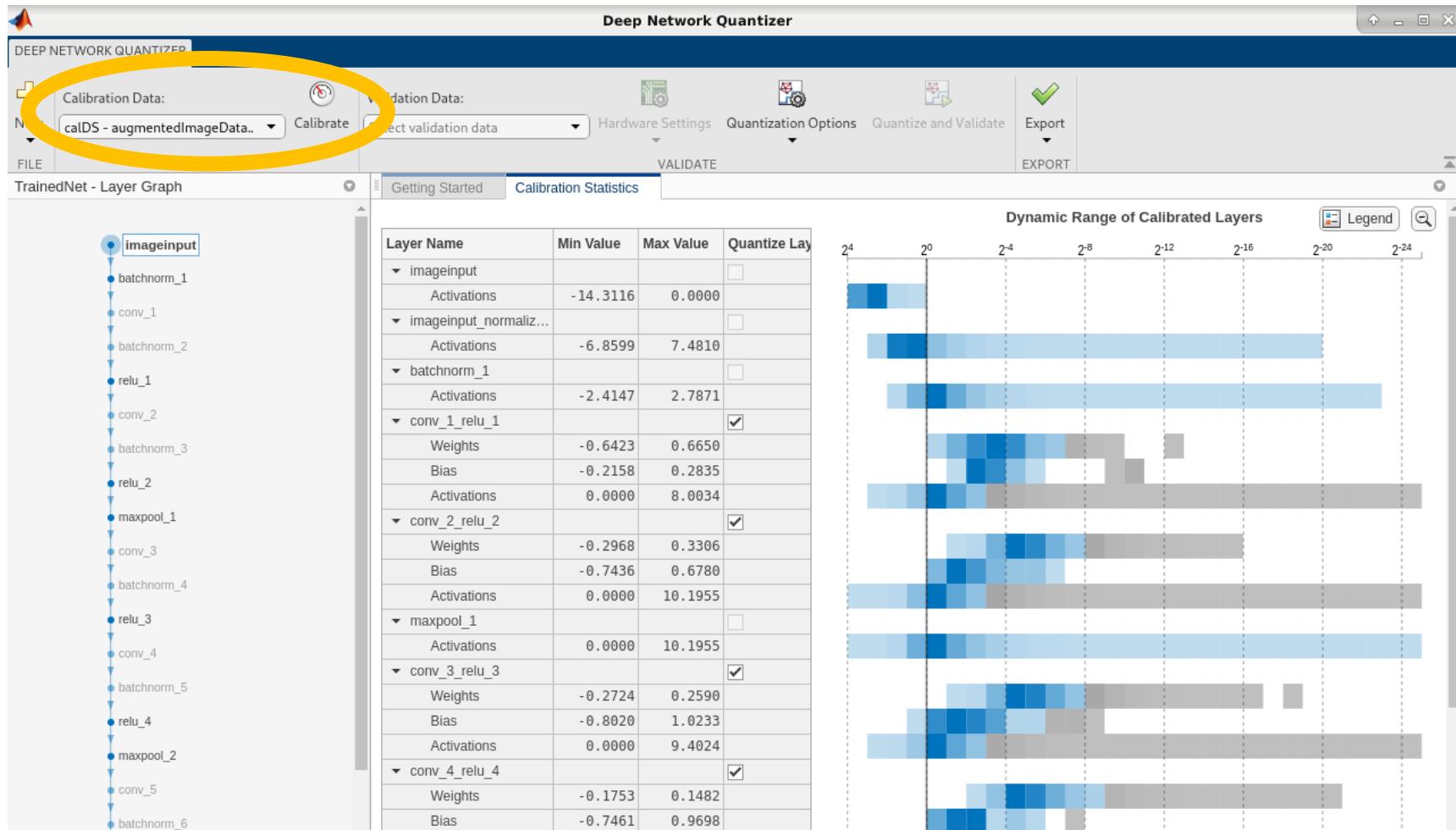


Use Deep Network Quantizer to Optimize the Inference Network

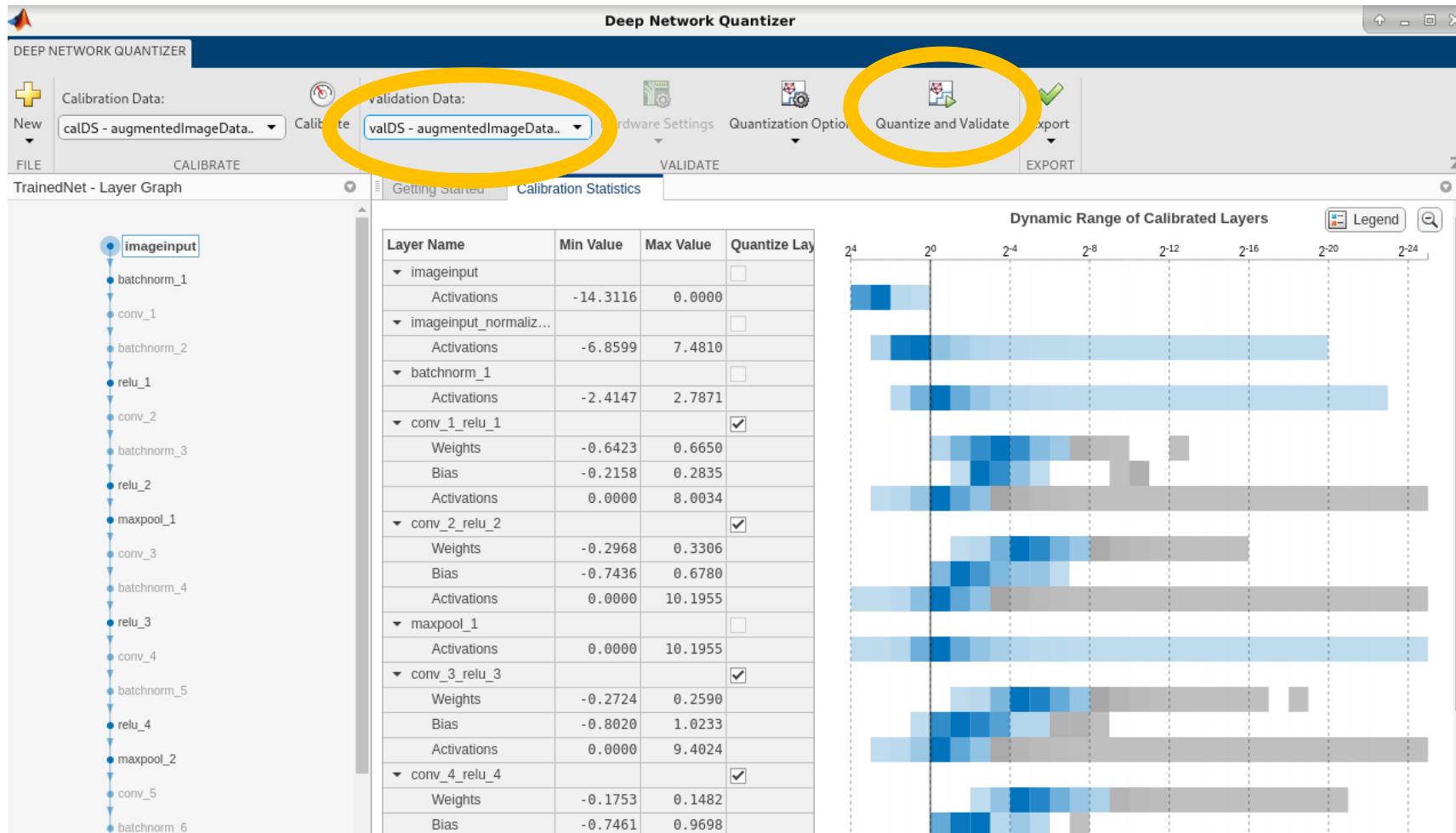
```
1 load('trainedNet');
2 analyzeNetwork(trainedNet);
3 numData = size(xTrain);
4 numData = numData(end);
5 augImds = augmentedImageDatastore(trainedNet.Layers(1).InputSize, xTrain, yTrain);
6 calDS = augImds.subset(1:floor(numData * 0.8));
7 valDS = augImds.subset(floor(numData * 0.8)+1:numData);
8 dq = dlquantizer(trainedNet, 'ExecutionEnvironment', 'GPU');
9 dq.calibrate(calDS)
```

- 加载预训练网络
- 创建数据集: 调校 – 80%, 验证 – 20%
- 运行 Deep Network Quantizer App

Deep Learning Quantization: 声音场景分类



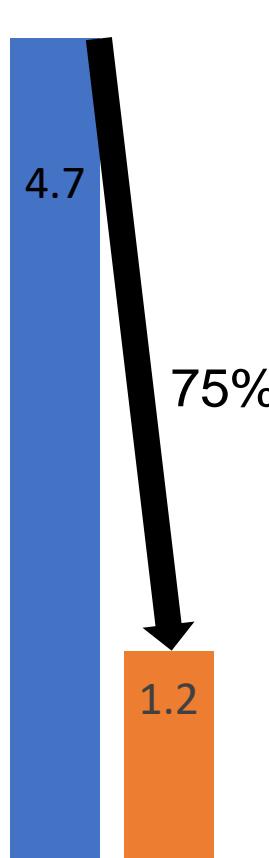
Deep Learning Quantization: 声音场景分类



Deep Learning Quantization: 声音场景分类

✓ Validation Results

Memory (MB)



Top-2 Accuracy

Learnable Parameters

■ FP32 ■ INT-8

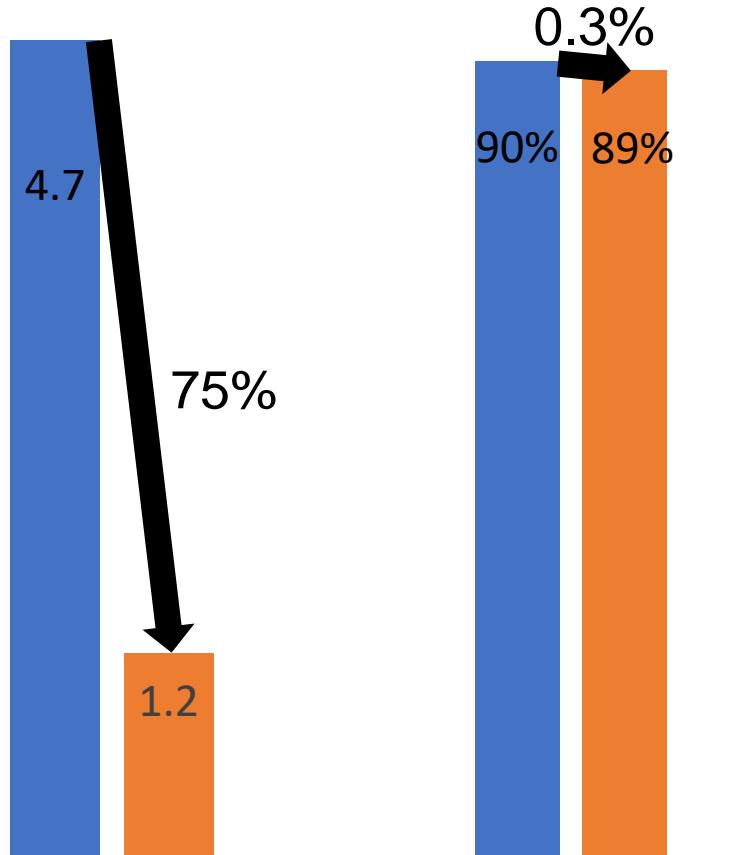
Top-2 Accuracy

■ FP32 ■ INT-8

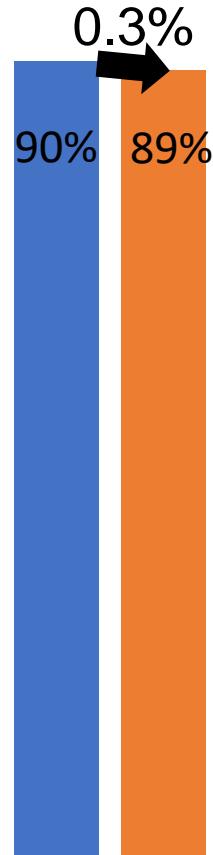
Deep Learning Quantization: 声音场景分类

✓ Validation Results

Memory (MB)



Top-2 Accuracy

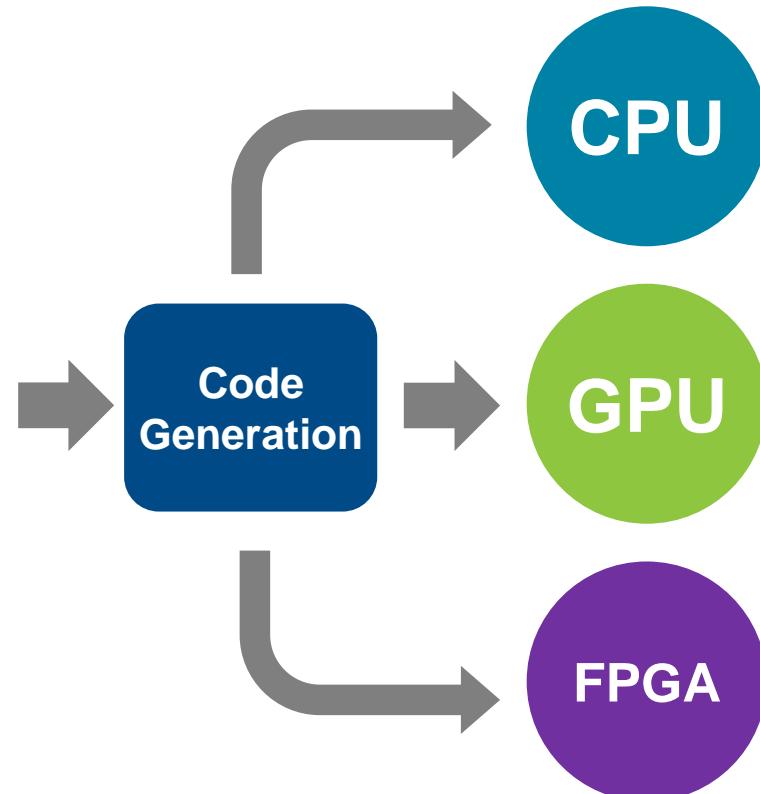
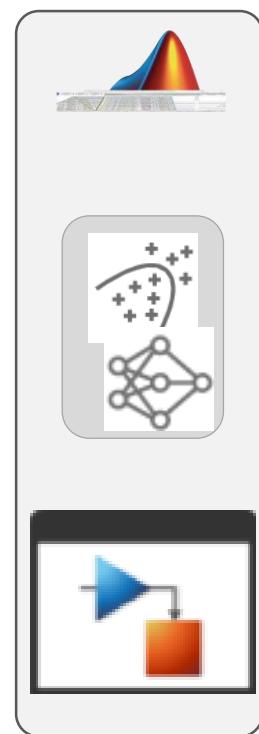


Learnable Parameters

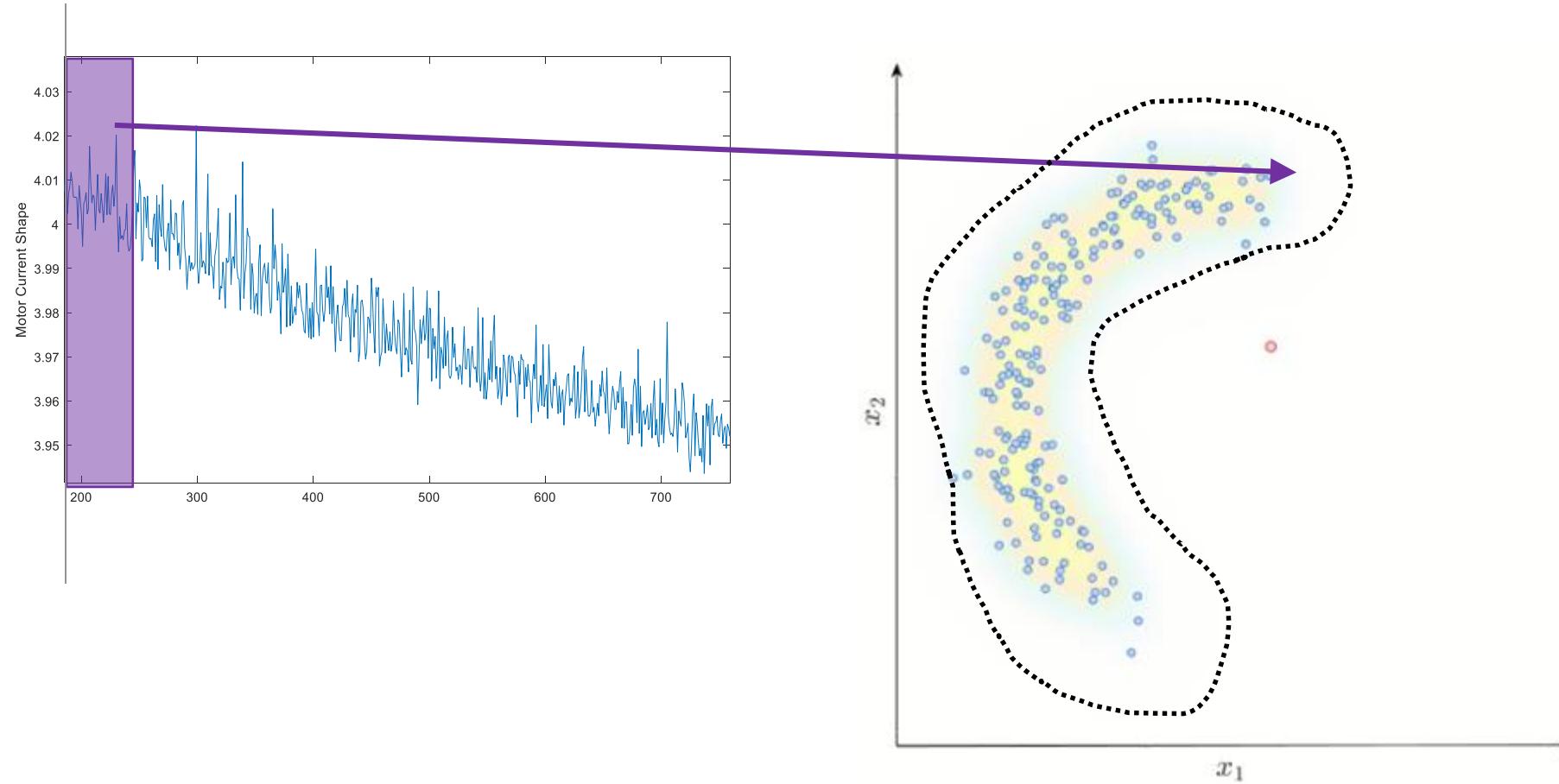
■ FP32 ■ INT-8

Top-2 Accuracy

■ FP32 ■ INT-8

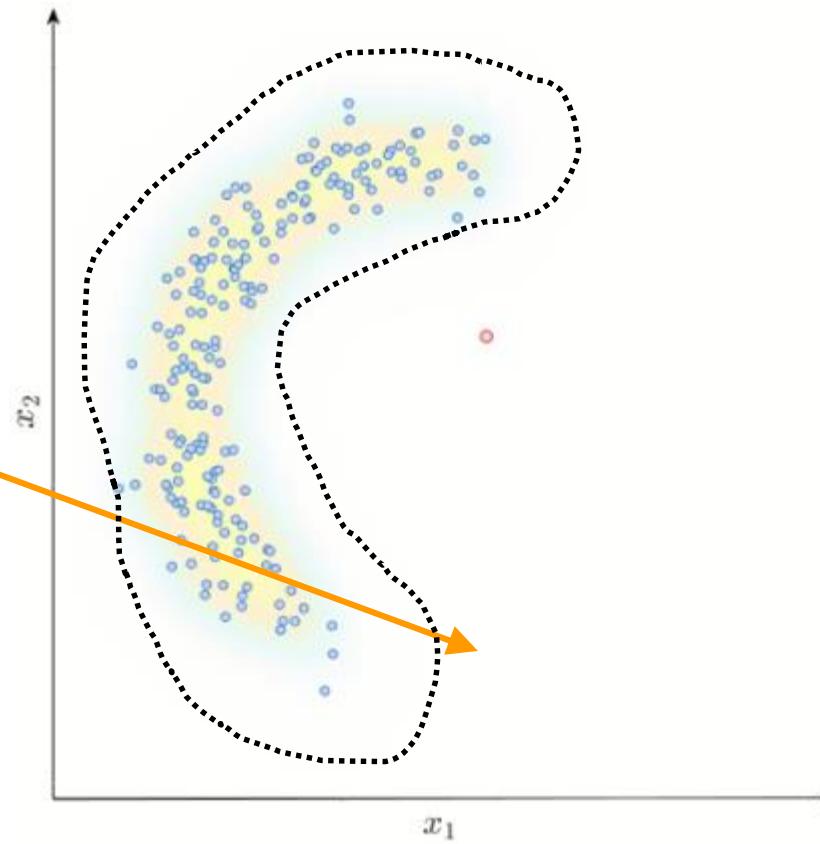
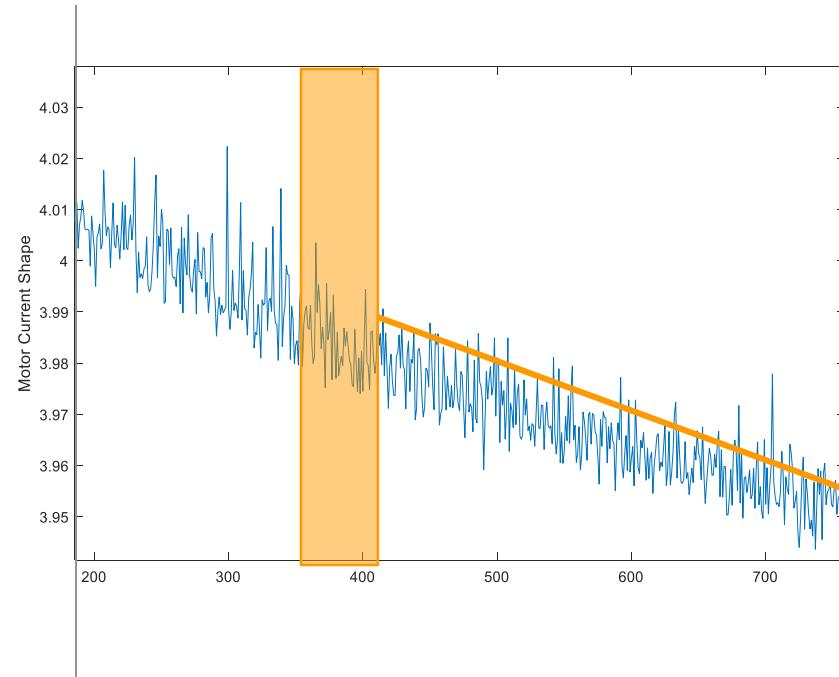


AI models 反应系统行为和环境

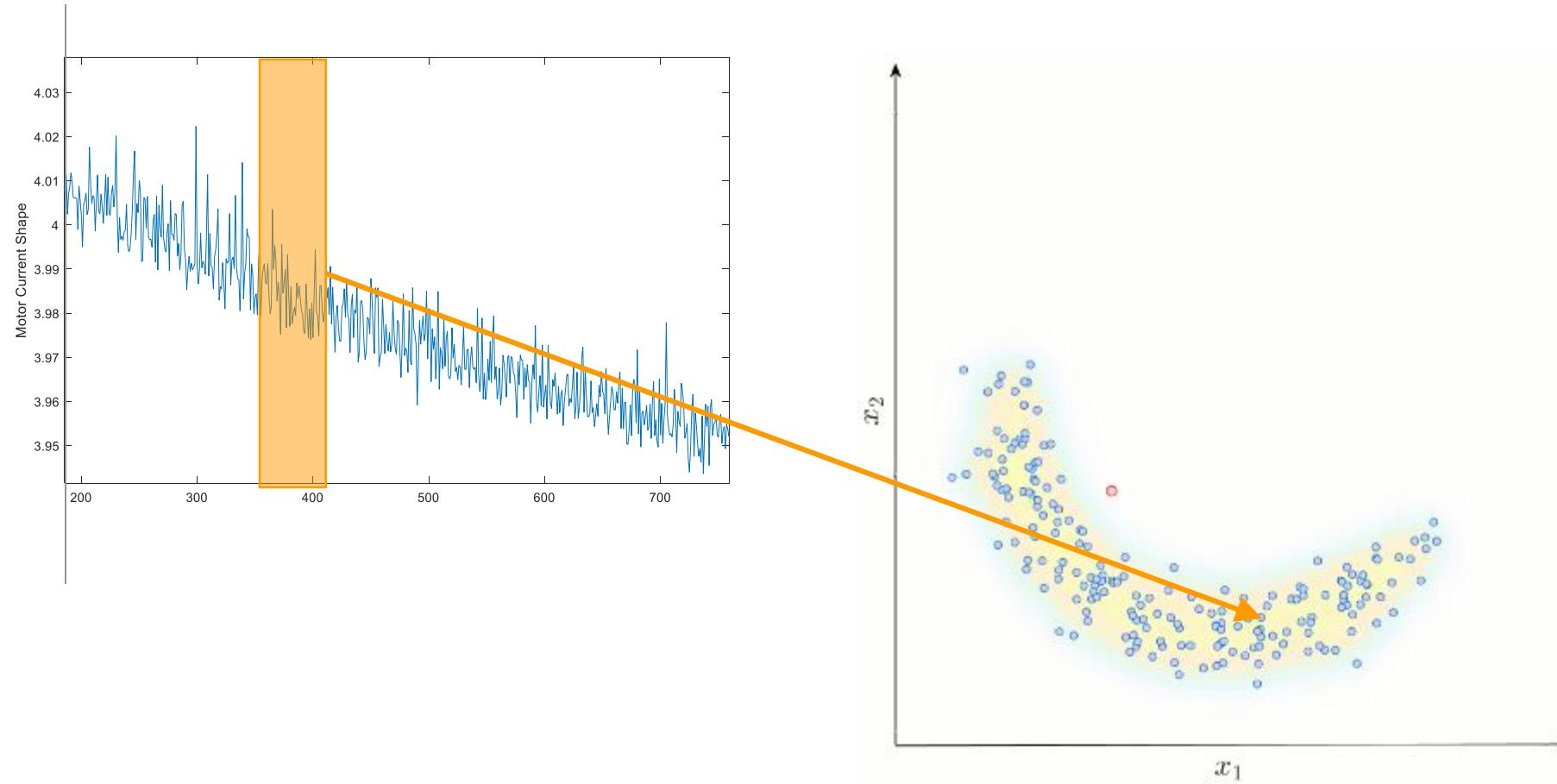


(illustration only; not based on actual data)

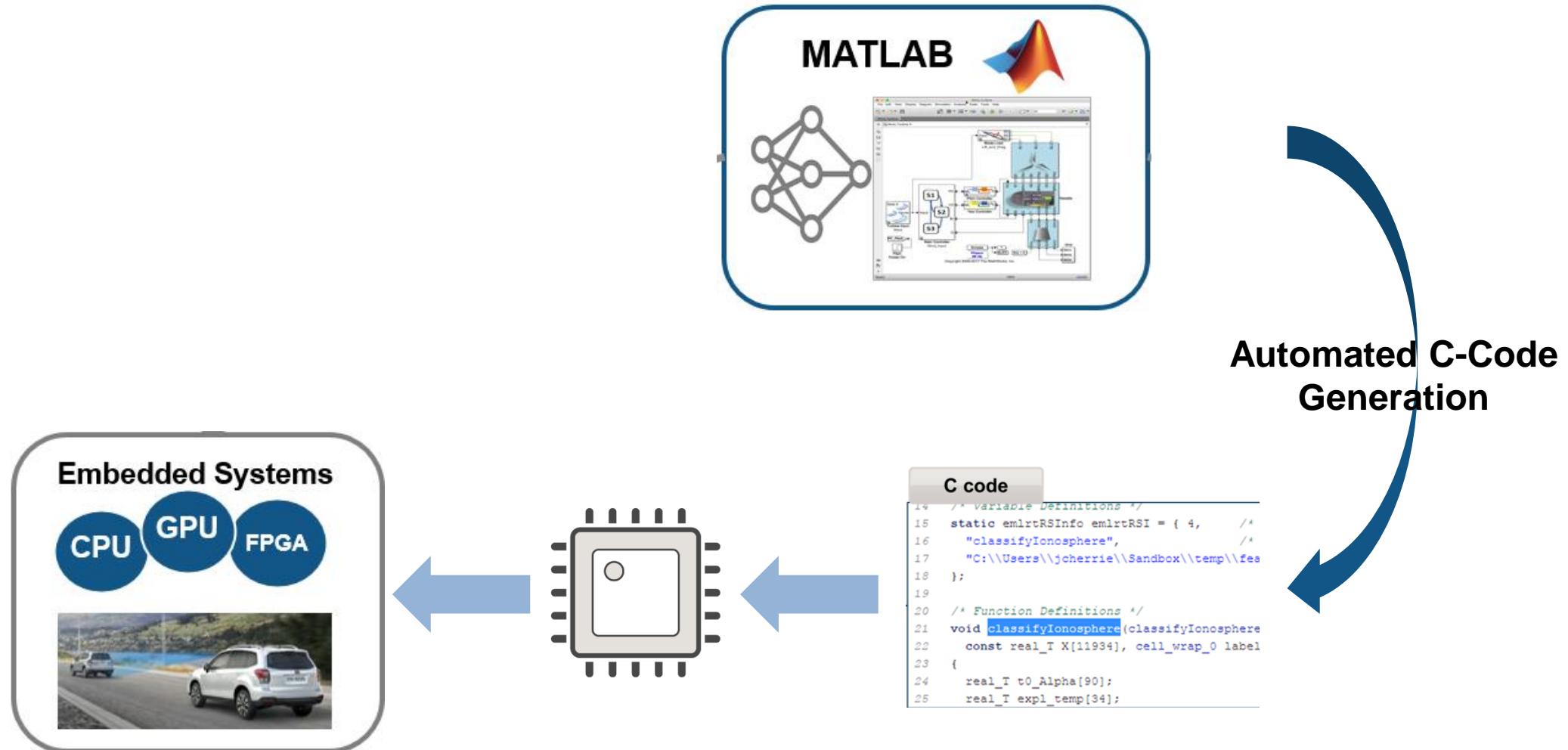
AI models反应系统行为和环境



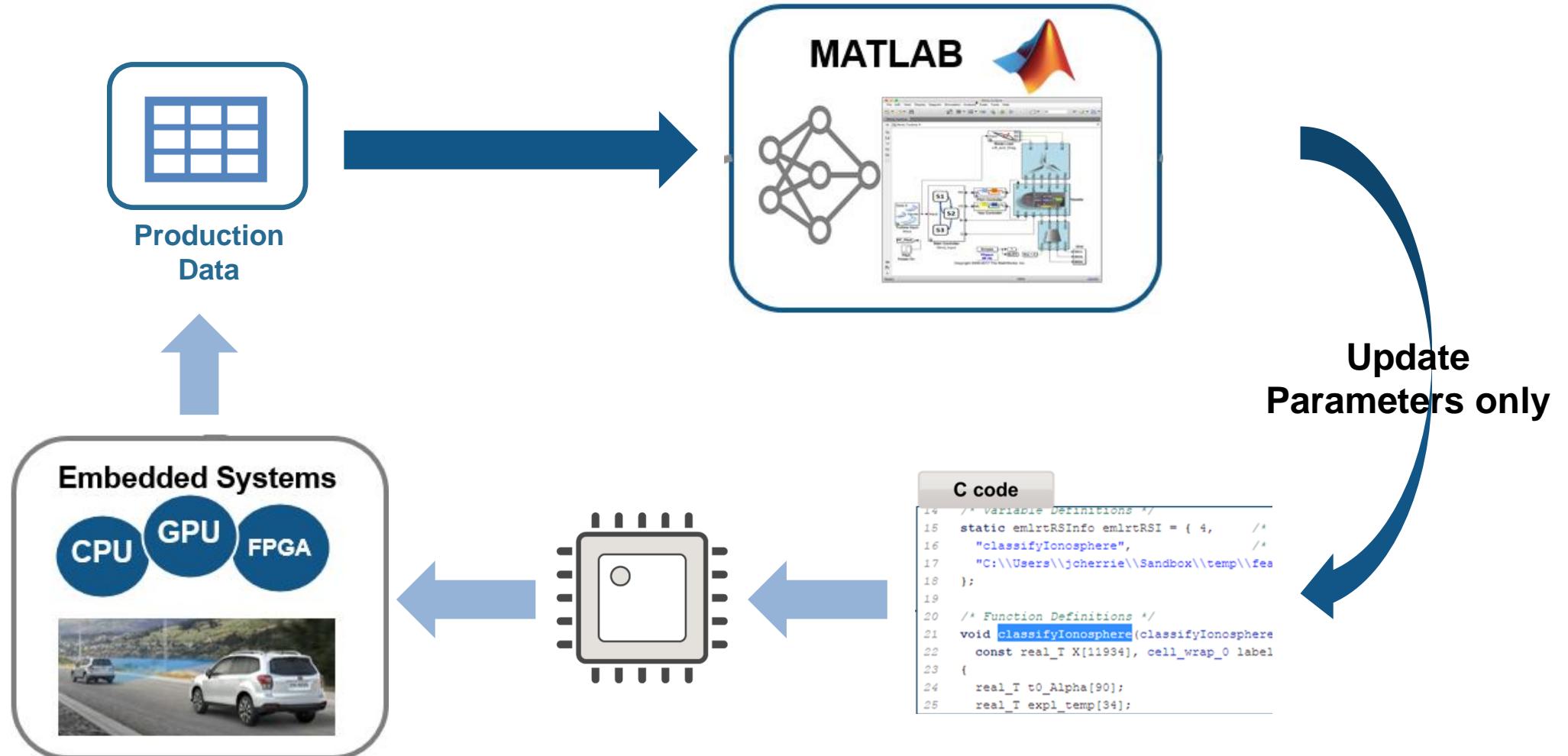
模型与算法需要自动修正



嵌入式系统中的模型更新



嵌入式系统中的模型更新



Agenda

算法和模型部署



三个特定问题:

1. 嵌入式硬件的约束
2. 环境和系统的持续变化
3. 扩展到企业生产环境





上海电气
SHANGHAI ELECTRIC

滑动轴承APP开发及应用部署

付玉敏, 上海电气集团中央研究院高级主管工程师



MATLAB EXPO 2021

Content 目录

01\ 项目背景

02\ 开发部署

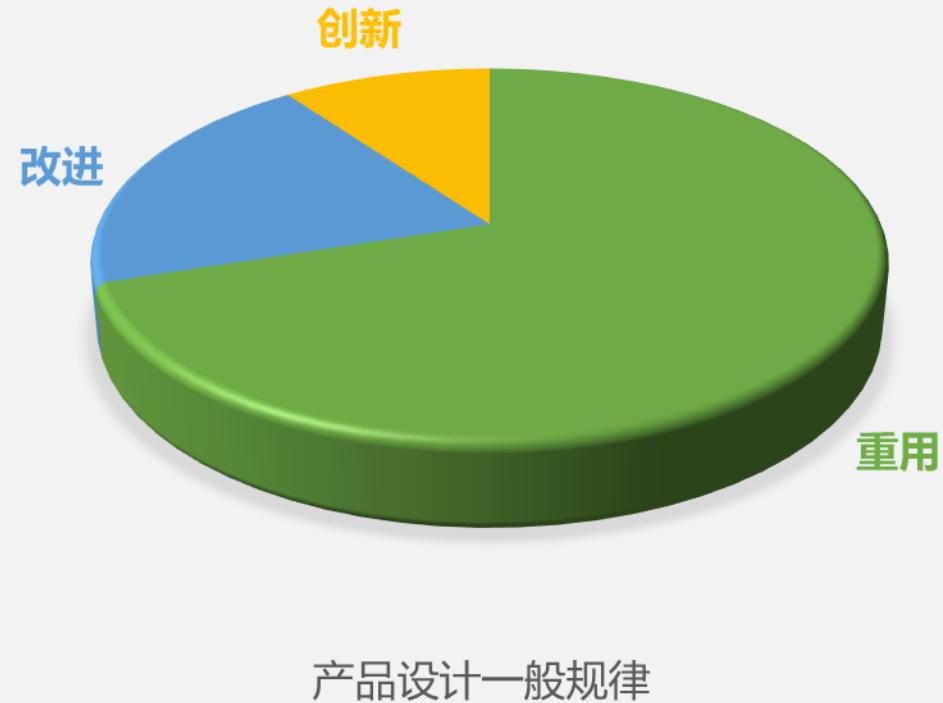
03\ 成果展示

Chapter 01

项目背景

项目背景

◆ 价值创新推动知识经验工具化发展



◆ 共享经济催生软件工具网络化趋势



项目背景

重要程度

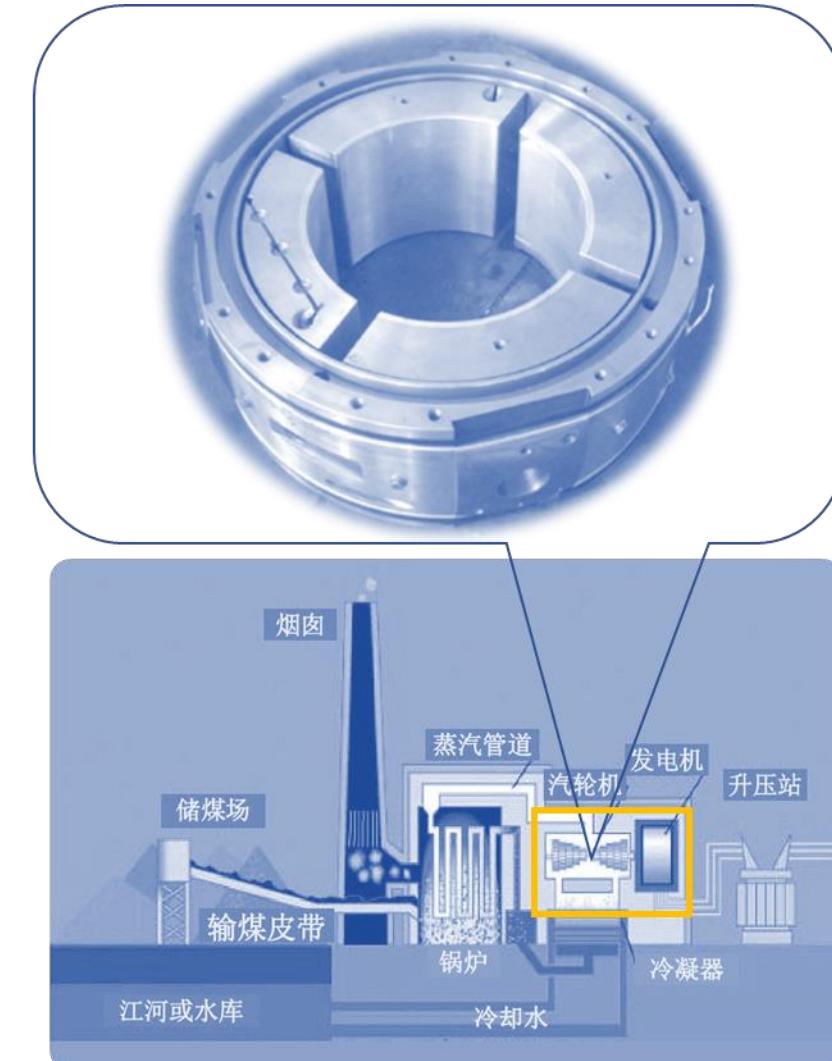
- 发电设备是电气主业之一
- 轴承是发电主设备的关键部件

研发创新

- 从逆向设计到自主研发
- 从依赖经验到依托数据

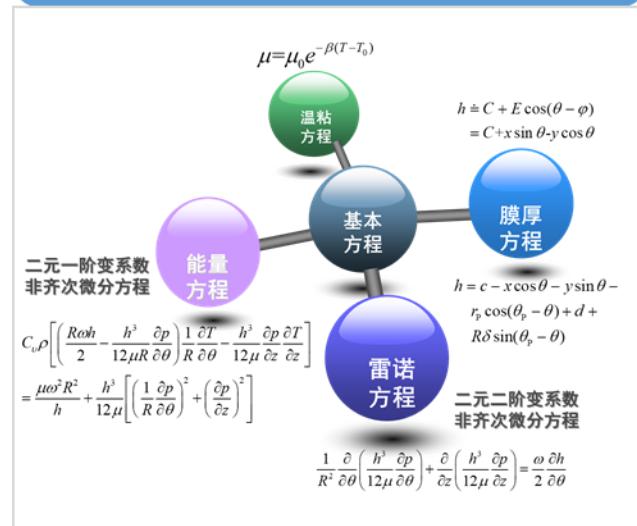
业务转型

- 从提供设备到提供设备+服务
- 发电设备运维服务平台建设

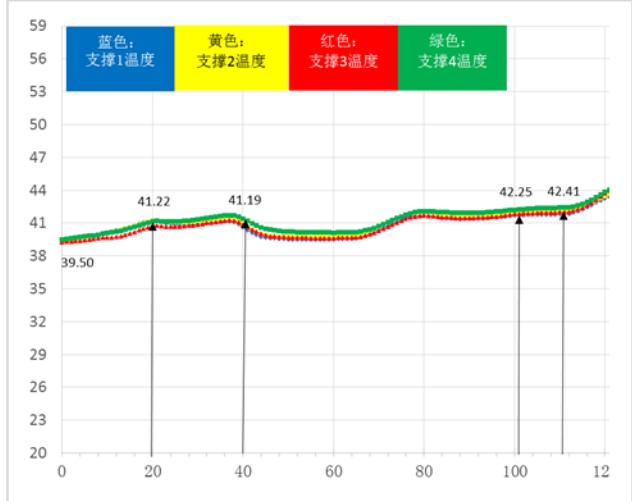


问题与挑战

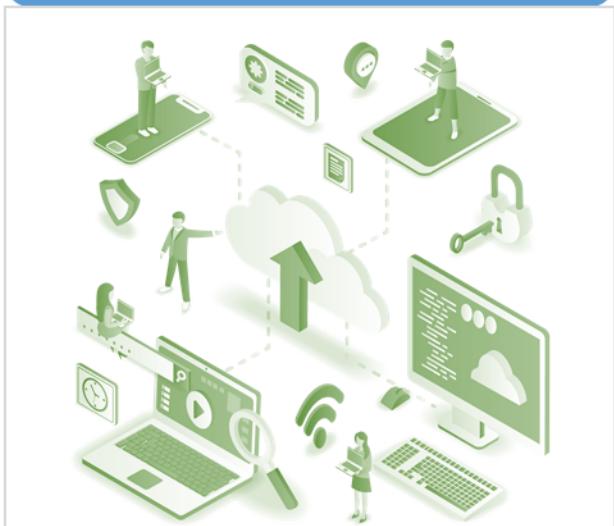
➤ 如何实现复杂机理模型工具化开发?



➤ 如何结合试验数据实现模型修正优化?



➤ 如何实现协同共享及数据安全?

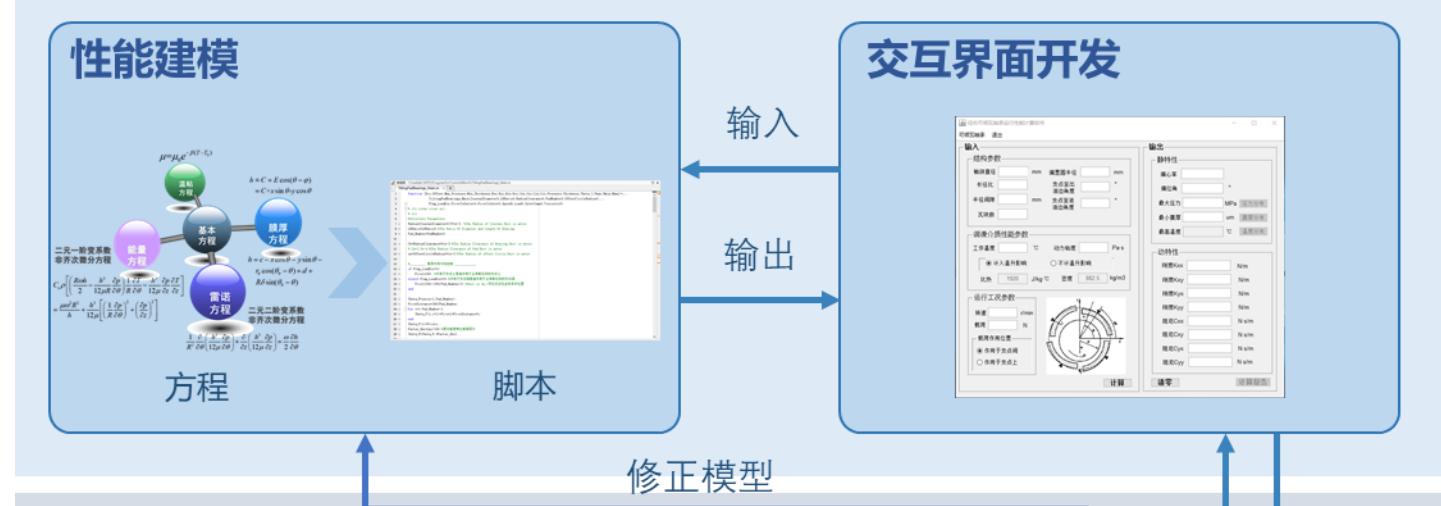


Chapter 02

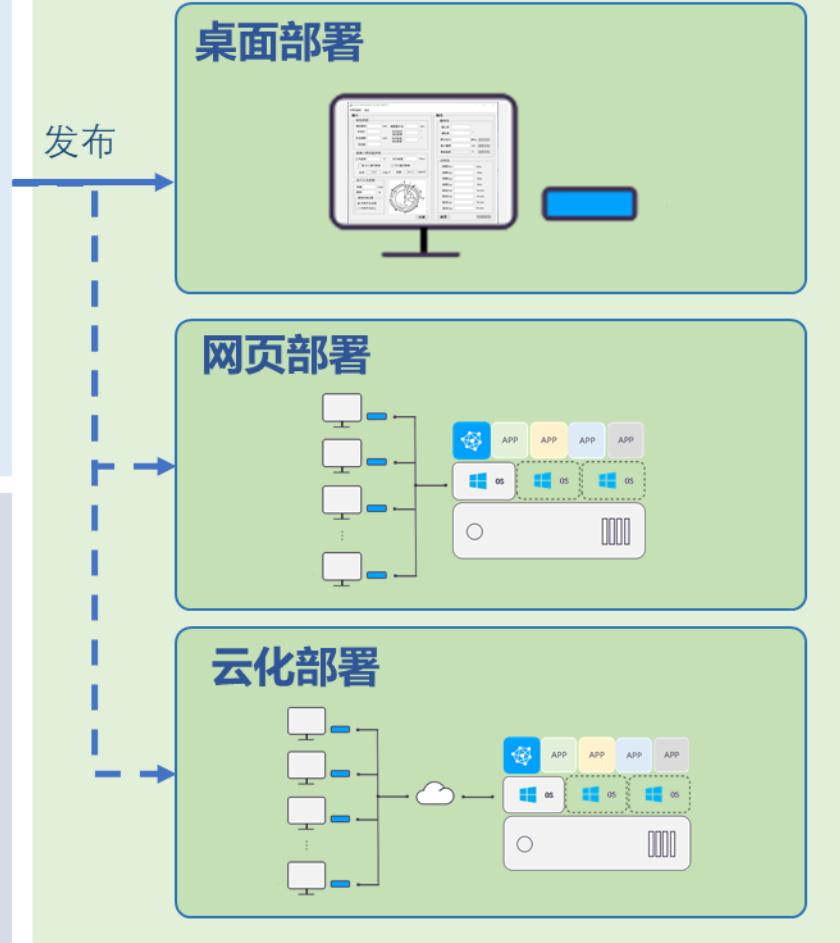
开发部署

工作流图

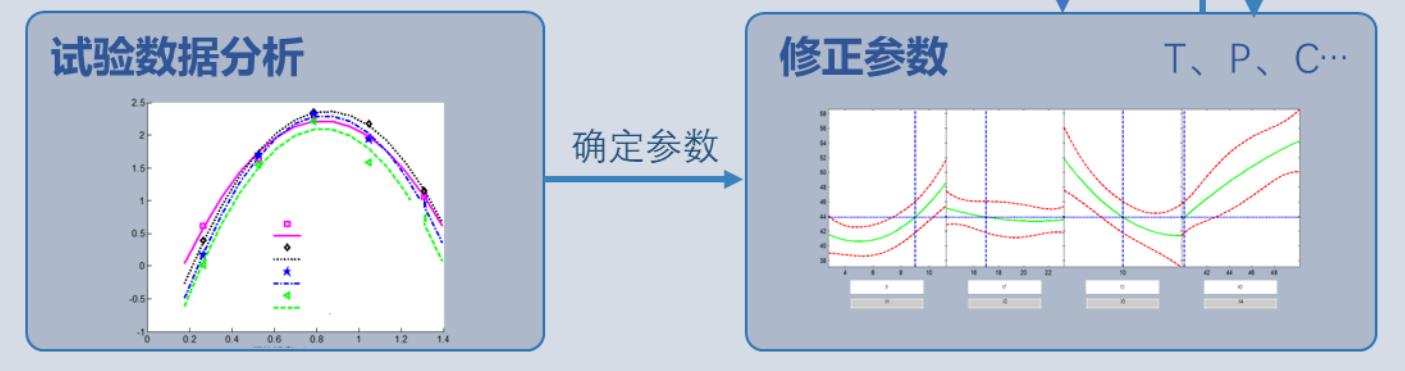
开发



部署

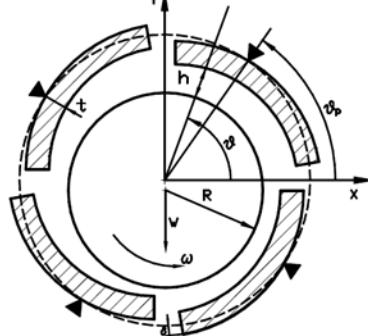


修正



性能建模

对象类型



基本方程

温粘方程: $\mu = \mu_0 e^{-\beta(T-T_0)}$

膜厚方程: $h = c - x \cos \theta - y \sin \theta - r_p \cos(\theta_p - \theta) + d + R \delta \sin(\theta_p - \theta)$

雷诺方程: $C_v \rho \left[\frac{R \partial h}{2} - \frac{h^3}{12 \mu R} \frac{\partial p}{\partial \theta} \right] \frac{1}{R} \frac{\partial T}{\partial \theta} - \frac{h^3}{12 \mu} \frac{\partial p}{\partial z}$

二元一阶变系数非齐次微分方程: $\frac{\mu \omega^2 R^2}{h} + \frac{h^3}{12 \mu} \left[\left(\frac{1}{R} \frac{\partial p}{\partial \theta} \right)^2 + \left(\frac{\partial p}{\partial z} \right)^2 \right]$

能量方程: $\frac{1}{R^2} \frac{\partial}{\partial \theta} \left(\frac{h^3}{12 \mu} \frac{\partial p}{\partial \theta} \right) + \frac{\partial}{\partial z} \left(\frac{h^3}{12 \mu} \frac{\partial p}{\partial z} \right) = \frac{\omega}{2} \frac{\partial h}{\partial \theta}$

模型开发

```

function [Ecc, Offset, Rad, Pressure, Rin, Thickness, Eccx, Eccy, Kyn, Epx, Cpx, Cpy, Cpx, Pressure, Thickness, Theta, I, Pass, Bain, Bain] = ...
    TiltingPadBearings_Main([Ecc, Eccx, Eccy, Kyn, Epx, Cpx, Cpy, Pressure, Thickness, Theta, I, Pass, Bain, Bain];...
    Pad_LedOutDir, PivotToOutlet, PivotToInlet, Speed, Load, InletTemp, Viscosity)
%
```

模型参数设置:

- Radius_JournalDiameter=2/1e-3; // 轴颈半径, 单位为米
- LDRatio=0.25*10; // 轴颈直径与轴长之比
- Pad_Number=PadNumber0;
- Theta=0;
- ClearanceRadius=1e-3; // 轴颈与瓦块间隙, 单位为米
- Pad_Radius=PadRadius0;
- rpxOffsetCircleRadius=1e-3; // 偏心圆半径, 单位为米
- ...

代码注释:

- 数据作用域控制: if Flag_LoadDir==1, Pivot=1; else Pivot=0; 将两个点设为主承瓦块的固定点。
- elseif Flag_LoadDir==0, 将两个点设为主承瓦块的转动点。
- Pivot=1->(2*Pad_Number)/2; 将点1设为主承瓦块在坐标系中位置。
- ...
- Theta_P=Theta(I,Pad_Number);
- PivotDistance=2*Pad_Number;
- for i=1:Pad_Number-1
 Theta_P(i+1)=Pivot1+PivotDistance*i;
end
- Theta_P(1)=Pivot1;
- Factor_Unit=pi/100; // 度与弧度单位转换因子
- Theta_P=Theta_P.*Factor_Unit

算法设计

雷诺方程:

$$\frac{\left(\frac{\bar{h}^3}{\bar{\mu}}\right)_{i+\frac{1}{2},j} \bar{p}_{i+1,j} - \left(\frac{\bar{h}^3}{\bar{\mu}}\right)_{i-\frac{1}{2},j} \bar{p}_{i-1,j} + \left(\frac{\bar{h}^3}{\bar{\mu}}\right)_{i+\frac{3}{2},j} \bar{p}_{i+3,j} + \left(\frac{\bar{h}^3}{\bar{\mu}}\right)_{i-\frac{3}{2},j} \bar{p}_{i-3,j}}{(\Delta \theta)^2} +$$

$$\frac{\left(\frac{\bar{h}^3}{\bar{\mu}}\right)_{i,j+\frac{1}{2}} \bar{p}_{i,j+1} - \left(\frac{\bar{h}^3}{\bar{\mu}}\right)_{i,j-\frac{1}{2}} \bar{p}_{i,j-1} + \left(\frac{\bar{h}^3}{\bar{\mu}}\right)_{i,j+\frac{3}{2}} \bar{p}_{i,j+3} + \left(\frac{\bar{h}^3}{\bar{\mu}}\right)_{i,j-\frac{3}{2}} \bar{p}_{i,j-3}}{(\Delta z)^2} = \frac{\bar{h}_{i+\frac{1}{2},j} - \bar{h}_{i-\frac{1}{2},j}}{\Delta \theta}$$

差分格式:

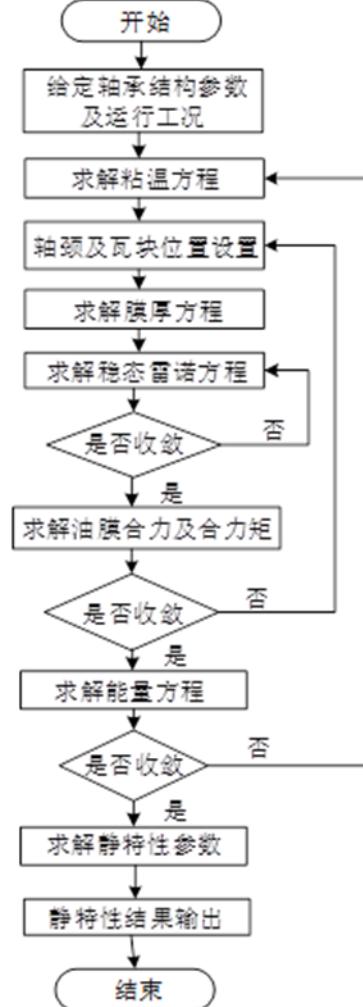
能量方程:

$$[\bar{h}_{i,j} - \frac{(\bar{h}_{i,j})^3}{\bar{\mu}_{i,j}} (\frac{\partial \bar{p}}{\partial \theta})_{i,j}] (\frac{\partial \bar{T}}{\partial \theta})_{i,j} - [\left(\frac{D}{L}\right)^2 (\bar{h}_{i,j})^3 (\frac{\partial \bar{p}}{\partial z})_{i,j}] (\frac{\partial \bar{T}}{\partial z})_{i,j} = T_{ca} \left[\frac{(\bar{h}_{i,j})^3}{3 \bar{h}_{i,j}} \left[\left(\frac{\partial \bar{p}}{\partial \theta} \right)_{i,j}^2 + \left(\frac{D}{L} \right)^2 \left(\frac{\partial \bar{p}}{\partial z} \right)_{i,j}^2 \right] \right]$$

偏微分方程:

代数方程组:

算法流程



模型修正

测试数据

主轴转速

工作负载

供油温度

供油压力

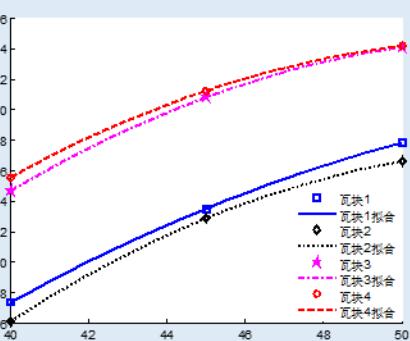
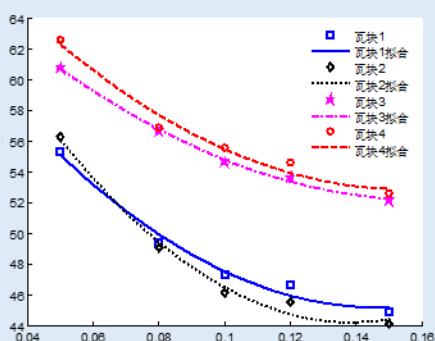
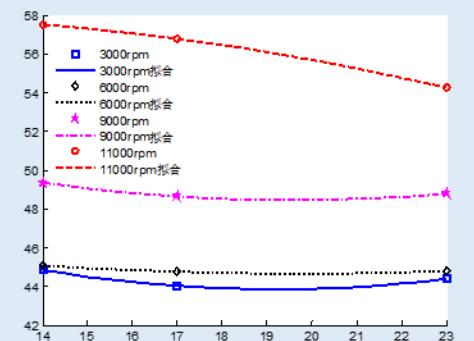
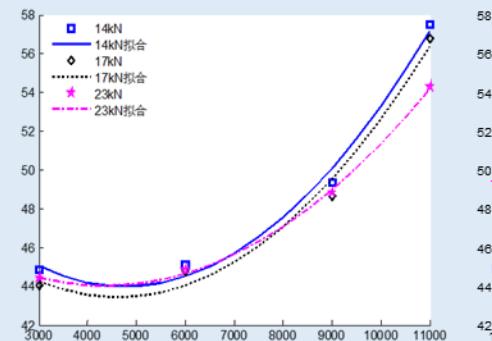
轴瓦温度

油膜压力

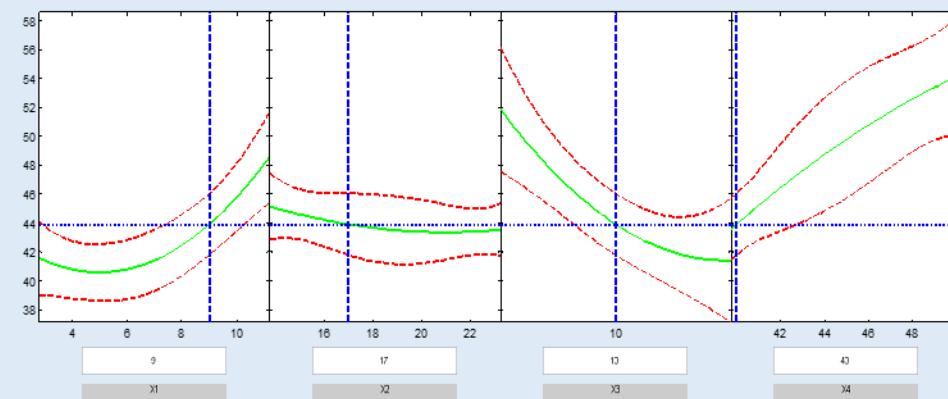
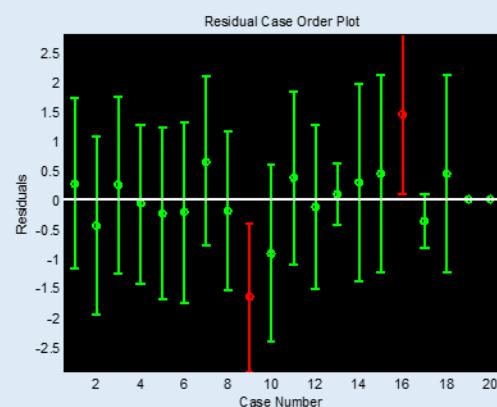
回油温度

.....

数据分析



多元二项式回归分析



交互界面开发

登录界面



主界面

- ✓ 计算参数输入
- ✓ 温升影响设置
- ✓ 载荷方向设置
- ✓ 计算参数输出
- ✓ 分布云图输出
- ✓ 计算报告生成



修正界面



边界修正

输入

工况个数	20		
参数个数	8		
导入测试数据			

	1	2	3	4
1	53.4600	47.8800	56.4500	57.
2	44.7300	43.1000	50.2600	49.
3	42.1000	41.2000	47.4900	45.
4	42.6300	42.5200	47.5000	44.
5	51.3000	46.0200	55.5000	56.
6	43.7000	41.9700	50.1400	48.
7	41.2300	40.2500	47.5300	44.
8	41.1400	41.0500	46.9700	44.
9	48.5100	44.0300	54.3400	54.
10	42.7700	40.7100	50.5800	48.
11	40.7100	39.4100	48.4000	44.
12	40.8100	40.5500	48.2000	44.
13	55.3300	56.3100	60.7900	62.
14	49.4500	49.0700	56.6500	56.
15	47.3700	46.1300	54.7000	55.

输出

	1	2	3	4
1	-21.7114	-94.9463	-82.4582	-90.4°
2	-2.0721	-2.5959	-1.4098	-2.26
3	-1.4395	-1.3822	-0.8394	-1.11
4	-2.9696	-5.0170	-2.3886	-3.11
5	3.7661	7.5408	6.2471	6.91
6	0.2076	0.2291	0.1581	0.29
7	0.0343	0.0319	0.0249	0.01
8	0.0987	0.1922	0.0770	0.11

回归系数:

	1	2	3	4
1	0.9940	0.9858	0.9984	0.9
2	165.3366	69.1842	628.1062	237.0
3	4.5065e-08	1.3937e-06	2.2259e-10	1.0784e-

回归模型检验:

	1	2	3	4
1	0.9940	0.9858	0.9984	0.9
2	165.3366	69.1842	628.1062	237.0
3	4.5065e-08	1.3937e-06	2.2259e-10	1.0784e-

结论:

修正成功!

返回主界面

应用部署

桌面部署



网页部署



云化部署



典型特征

- 独立桌面应用程序
- 执行不依赖MATLAB

应用范围

- 团队或部门内部

开发者&用户配置

- 开发者: MATLAB Complier
- 用户: MATLAB Runtime

优缺点

- 独立应用，可脱离MATLAB执行
- 一旦下发应用，不受开发者控制；版本更新发布不及时

- 程序在Web浏览器中运行
- 内网/局域网共享

- 团队、部门或企业内部

- 开发者: App Designer、MATLAB Complier、Web App compiler、MATLAB Runtime

- 通过浏览器访问应用程序
- 用户无需安装或拷贝任何文件，便于知识产权保护；版本及时更新发布
- 共享范围有限

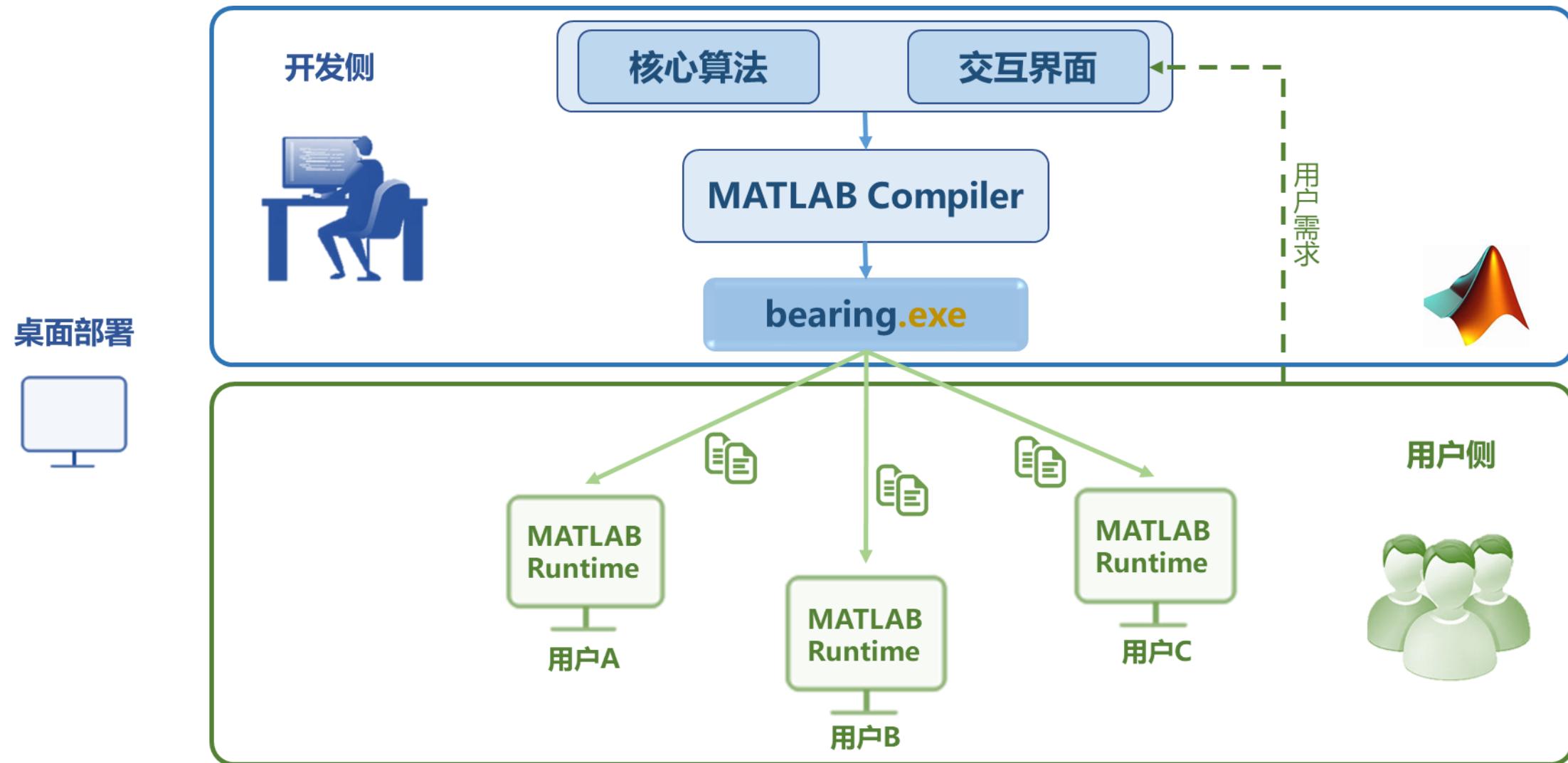
- 程序在Web浏览器中运行
- 公网共享

- 团队、部门、企业以及生态圈

- 开发者: App Designer、MATLAB Complier、Web App compiler、MATLAB Production Server、MATLAB Runtime

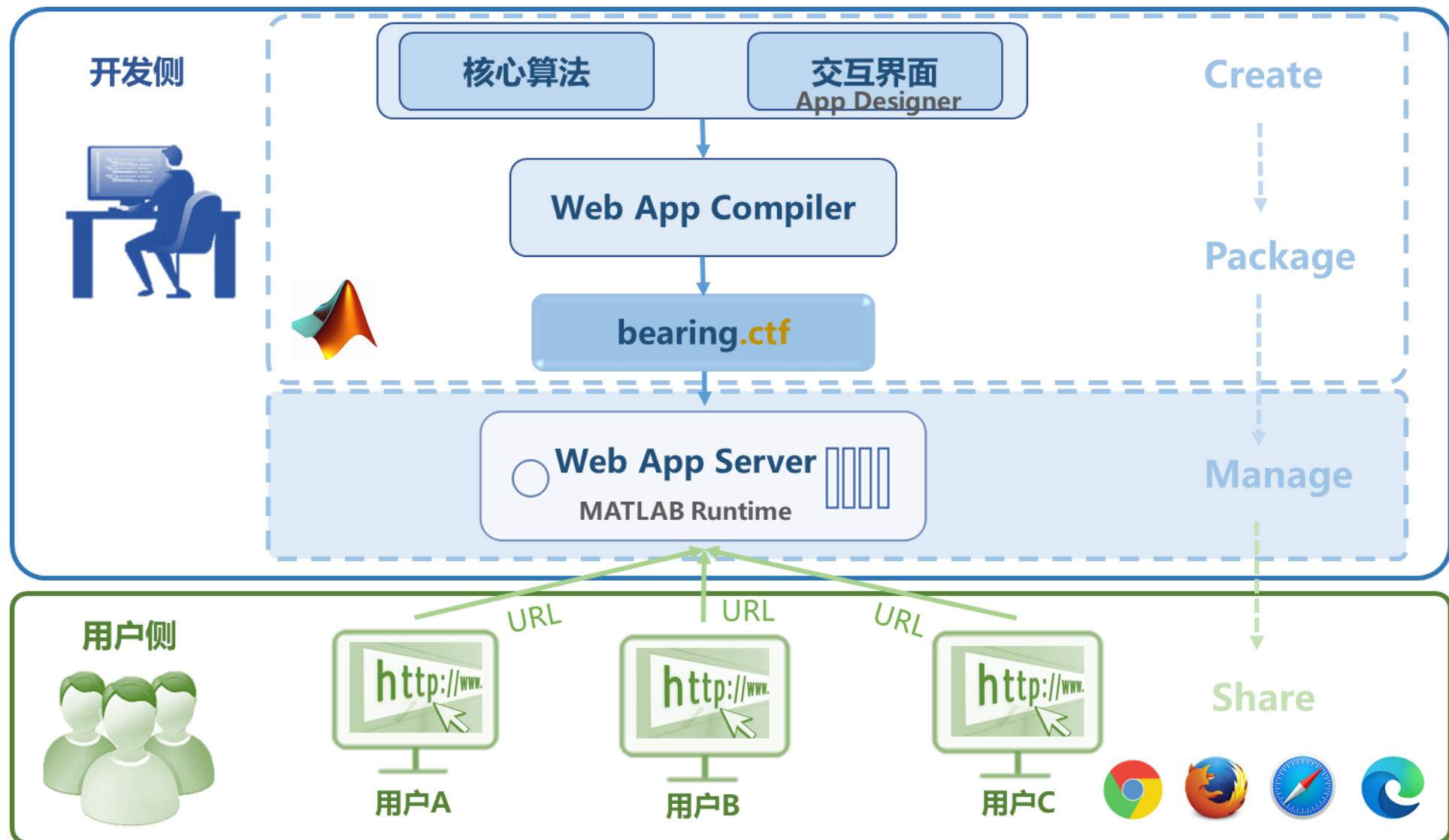
- 通过浏览器访问应用程序
- 用户无需安装或拷贝任何文件，便于知识产权保护；版本及时更新发布
- 大范围协同共享

应用部署



应用部署

网页部署



应用部署

云化部署



开发侧



- ✓ 云端部署
- ✓ 高并发访问
- ✓ 服务器可扩展

核心算法

MATLAB Compiler SDK

*bearing.ctf

MATLAB
Production Server
MATLAB Runtime (多版本)

Create

Package

Manage

用户侧



交互界面

独立

Web应用

可视化

第三方仪表板

集成

企业应用

Web API

Web API

Web API

Share



Chapter 03

成果展示

成果展示

开发工具典型页面



径向可倾瓦轴承运行性能计算软件

可倾瓦轴承 退出

输入

结构参数

轴颈直径	mm	偏置圆半径	mm
长径比		支点至出油边角度	°
半径间隙	mm	支点至进油边角度	°
瓦块数			

润滑介质性能参数

工作温度	℃	动力粘度	Pa·s
<input type="radio"/> 计入温升影响		<input checked="" type="radio"/> 不计温升影响	
比热	J/kg·℃	密度	kg/m ³

运行工况参数

转速	r/min
载荷	N
载荷作用位置	
<input checked="" type="radio"/> 作用于支点间	
<input type="radio"/> 作用于支点上	

输出

静特性

偏心率		
偏位角		
最大压力	MPa	压力分布
最小膜厚	um	膜厚分布
最高温度	℃	温度分布

动特性

刚度Kxx	N/m
刚度Kxy	N/m
刚度Kyx	N/m
刚度Kyy	N/m
阻尼Cxx	N·s/m
阻尼Cxy	N·s/m
阻尼Cyx	N·s/m
阻尼Cyy	N·s/m

计算

清零

计算报告

输出

回归系数:

	1	2	3	4
1	-21.7114	-94.9463	-82.4582	-90.41
2	-2.0721	-2.5959	-1.4098	-2.26
3	-1.4395	-1.3822	-0.8394	-1.11
4	-2.9696	-5.0170	-2.3886	-3.11
5	3.7661	7.5408	6.2471	6.90
6	0.2076	0.2291	0.1581	0.25
7	0.0343	0.0319	0.0249	0.02
8	0.0987	0.1922	0.0770	0.11

回归模型检验:

	1	2	3	4
1	0.9940	0.9858	0.9984	0.9
2	165.3366	69.1842	628.1062	237.0
3	4.5065e-08	1.3937e-06	2.2259e-10	1.0784e-07

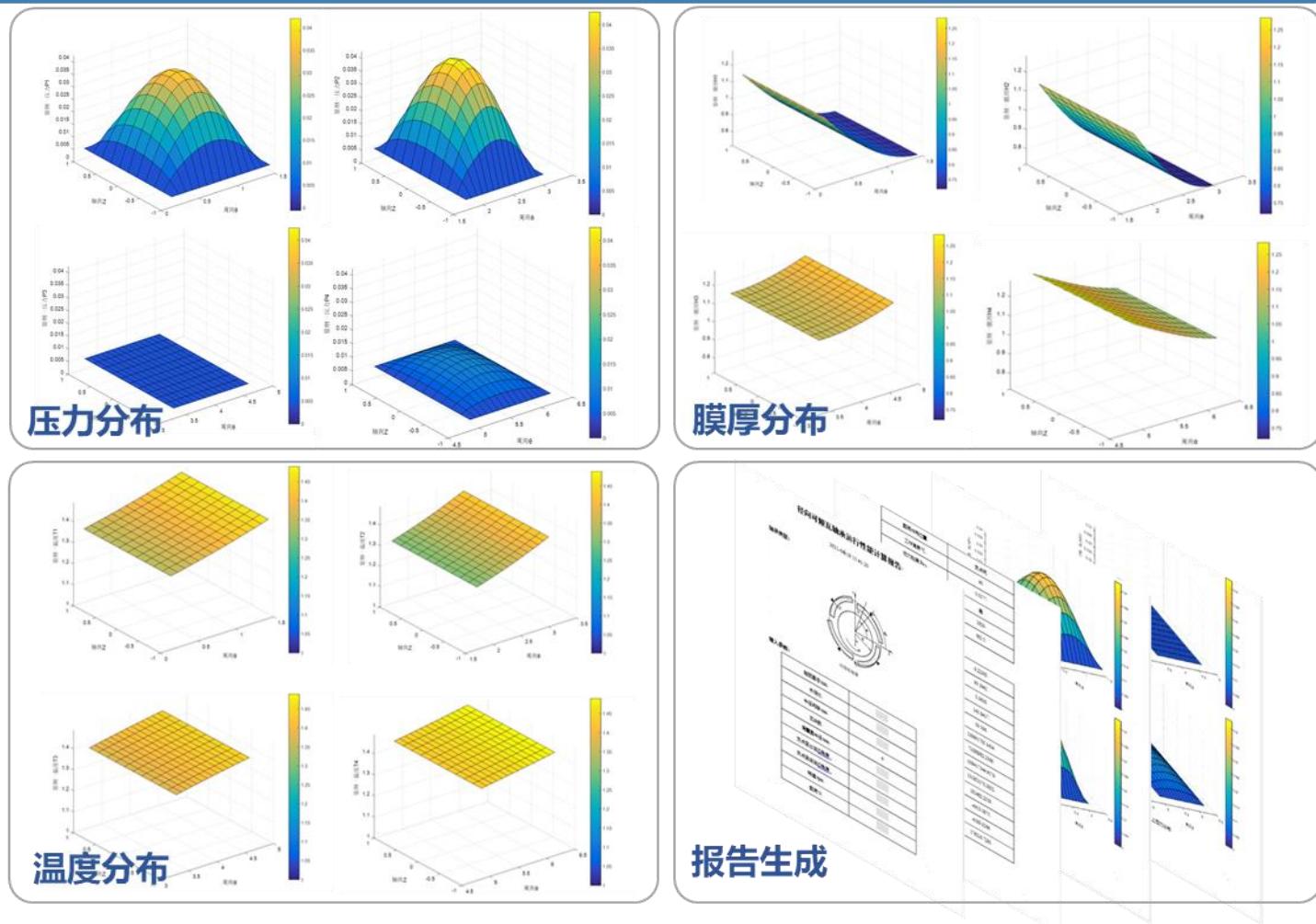
结论:

修正成功!

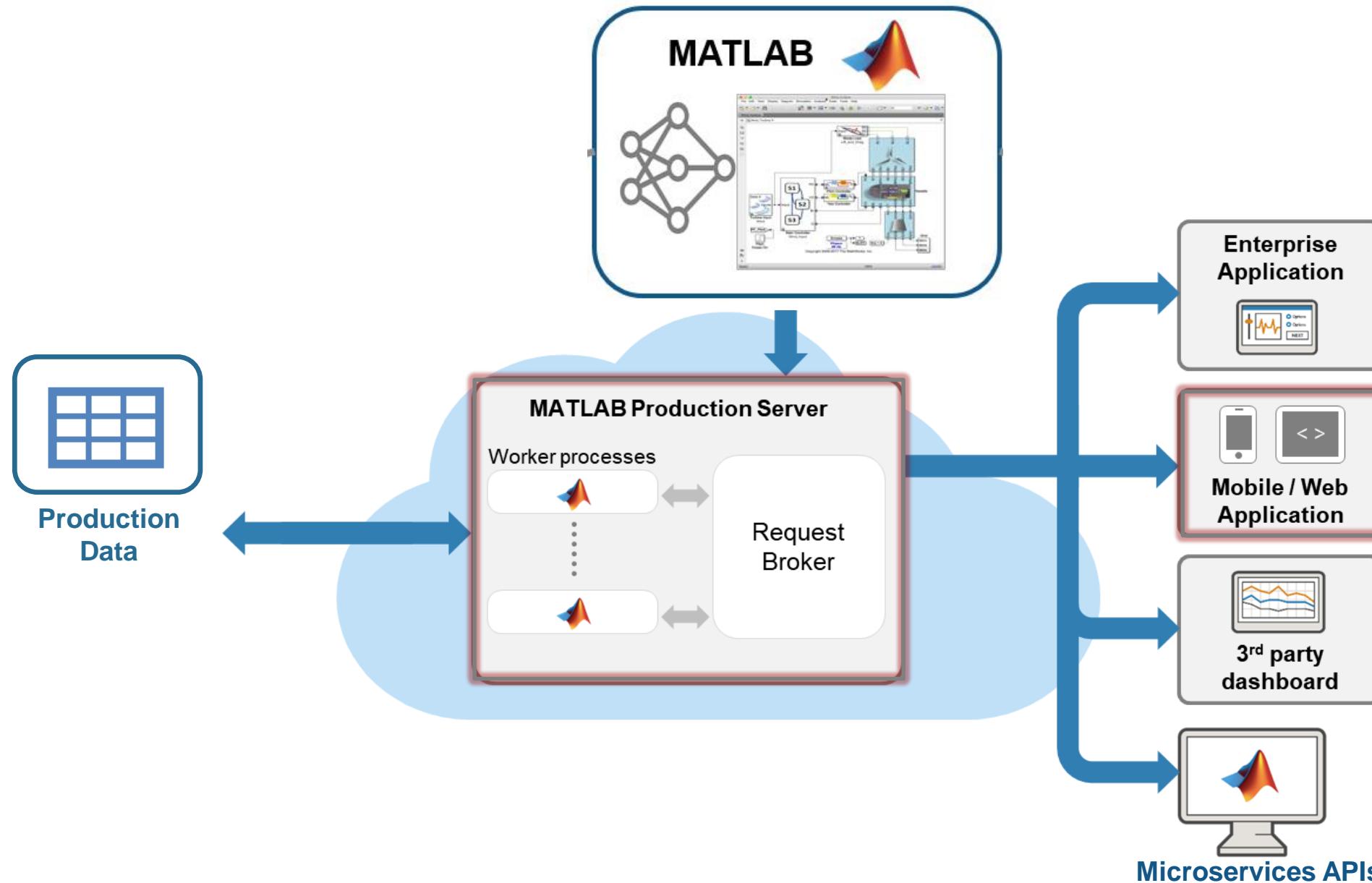
返回主界面

成果展示

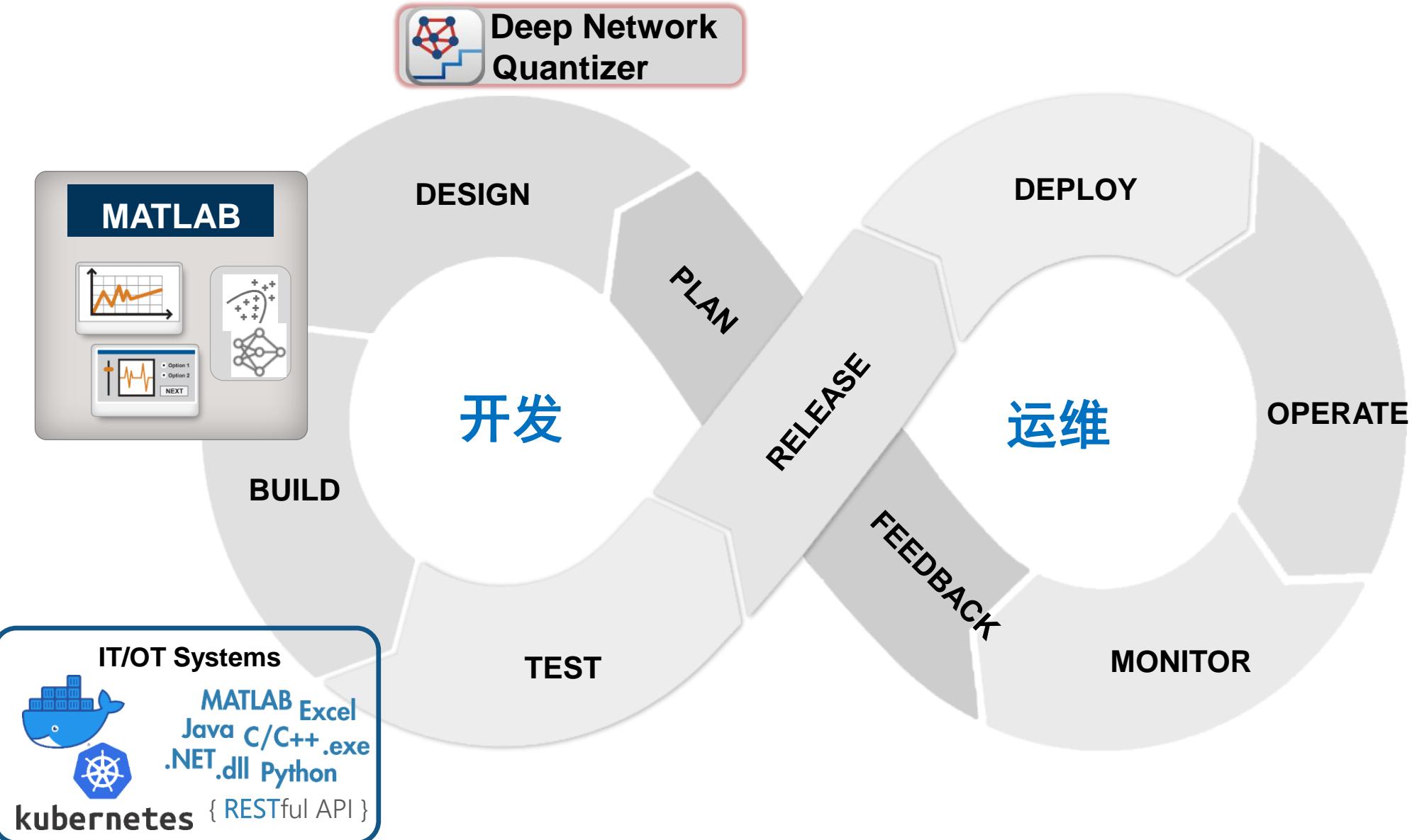
典型计算结果示例



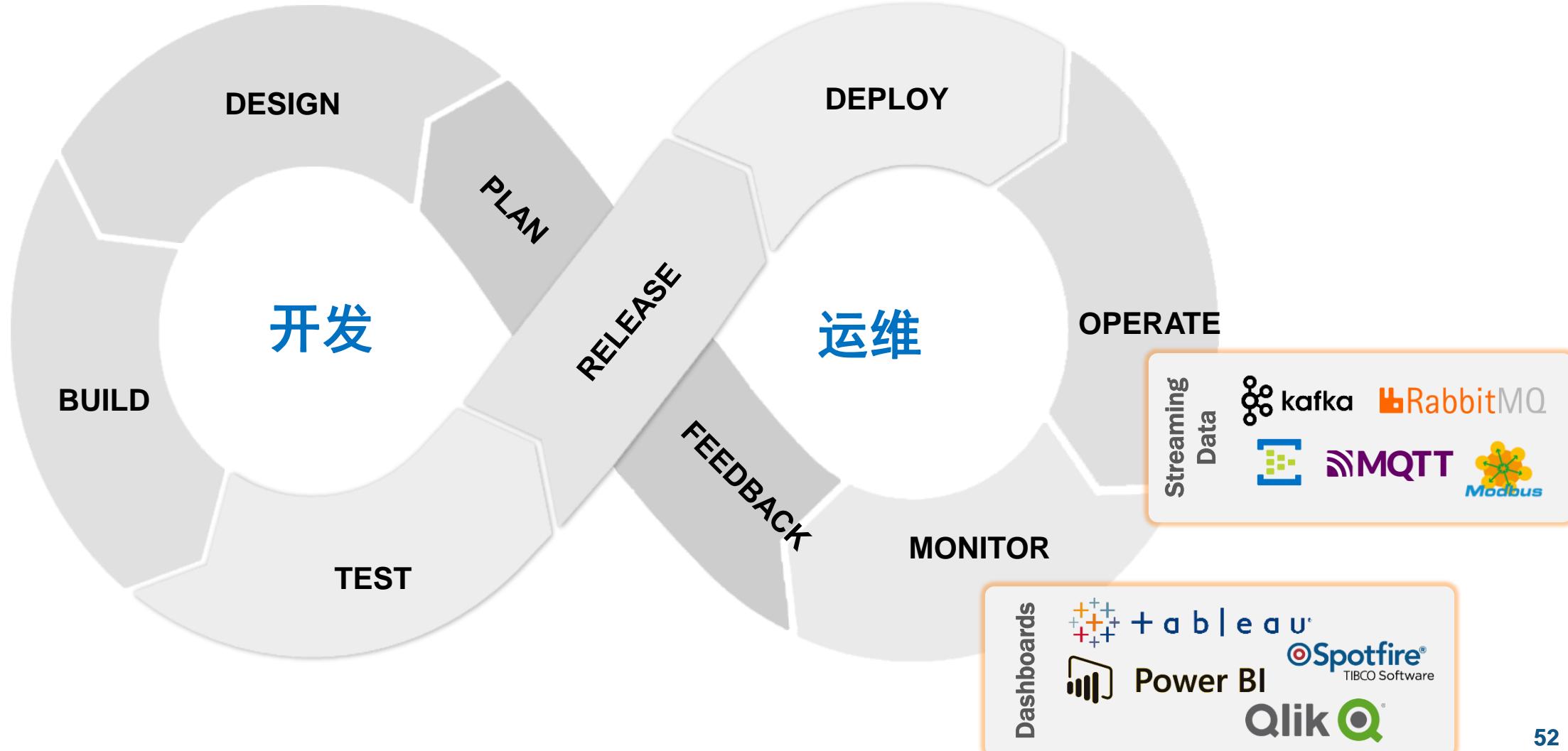
MATLAB Production Server与企业系统集成并扩展到生产系统



模型开发运维



模型开发运维



总结

从一个代码库部署到嵌入式和企业系统

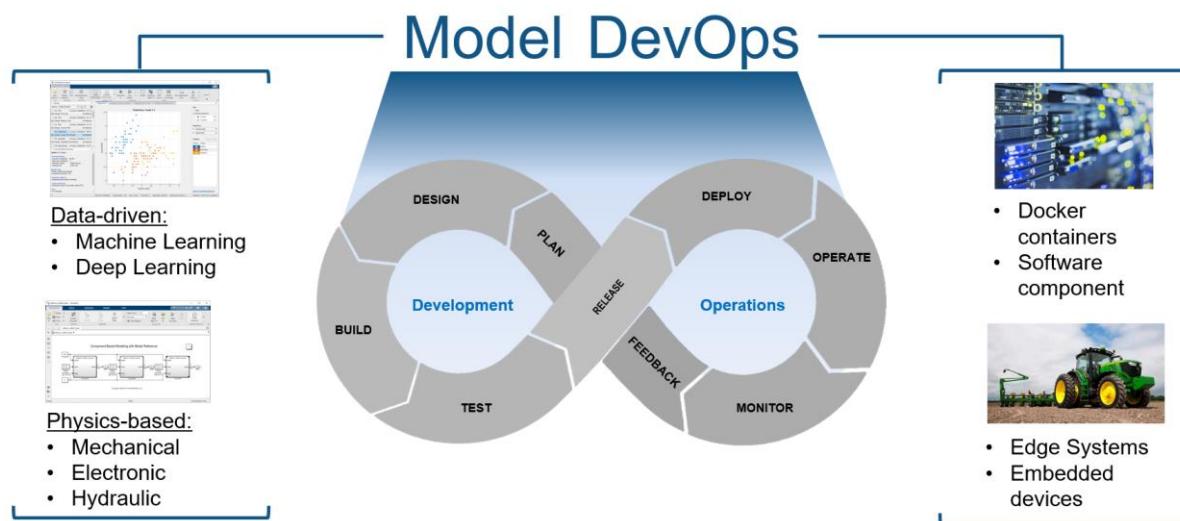
针对部署的挑战的工具：

- 通过量化/定点转换将模型调整适合于嵌入式硬件
- 增量地调整已部署的模型以保证性能
- 将算法与模型部署到生产系统

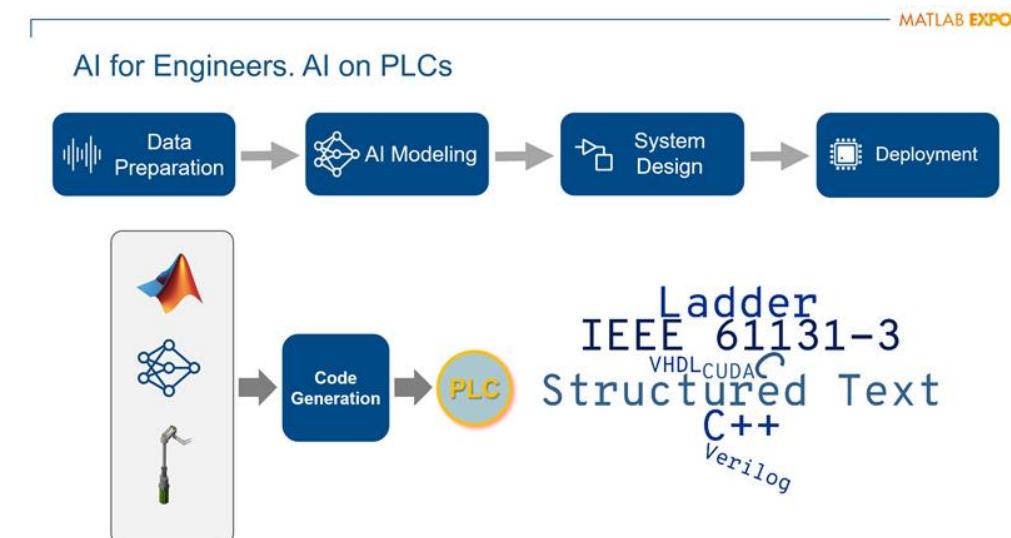
在一个框架内设计、部署和维护模型与算法

更多

查看我们的更多用户案例, 文档和示例并在MATLAB Online中尝试



*DevOps for Software and Systems:
Operationalization of Algorithms and Models*



Deploying AI on PLCs

MATLAB EXPO

2021

Thank you



© 2021 The MathWorks, Inc. MATLAB and Simulink are registered trademarks of The MathWorks, Inc. See mathworks.com/trademarks for a list of additional trademarks. Other product or brand names may be trademarks or registered trademarks of their respective holders.

